JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 20, No. 5 MAY, 1949

In This Issue

Page

Triangulation of a Two-Dimensional Continuum for the Purpose of the Approximate Solution of Second-Order Partial Differential Equations L. Tasny-Tschiassny 419

On the Velocity-Dependent Characteristics of High Frequency Tubes Julian K. Knipp 425

Barkhausen Noise and Magnetic Amplifiers. I. Theory of Magnetic Amplifiers J. A. Krumhansl and R. T. Beyer 432

On the Infinitely Long Cylindrical Antenna Charles H. Papas 437

Phase Contrast in Electron Microscope Images E. G. Ramberg 441

Flow through a Rocket Nozzle with and without Vibrational Equilibrium S. S. Penner 445

Note on the Substitution Principle for Steady Gas Flow R. C. Prim III 448

Stabilization of Breakdown Voltage and Measurement of Some Critical Currents in a Glow Tube A. B. Pal 451

High Frequency Total Emission Loading in Diodes Nicholas A. Begovich 457

Visual Method for Demonstrating the Path of Ultrasonic Waves through Thin Plates of Material Charles J. Burton and R. Bowling Barnes 462

High Current Carbon Arc and Its Mechanism Wolfgang Finkelnburg 468

Radio Doppler Investigation of Meteoric Heights and Velocities L. A. Manning, O. G. Villard, Jr., and A. M. Peterson 475

Letters to the Editor:

Improved Sectioning Technique for Electron Microscopy Richard F. Baker and Daniel G. Pease 480

Index to Advertisers xxxiv

BOARD OF EDITORS
Elmer Hutchisson, *Editor*

Associate Editors
R. Bowling Barnes
Otto Beeck
W. James Lyons
Philip Morrison
Paul E. Sabine
Raymond J. Seeger
Frederick Seitz
Charles S. Smith
J. R. Townsend

Publication of the American Institute of Physics

OFFICERS
George R. Harrison, *Chairman*
Wallace Waterfall, *Secretary*
George B. Pegram, *Treasurer*
John T. Tate, *Adviser on Publications*

GOVERNING BOARD
Luis W. Alvarez
J. W. Beams
Wallace R. Brode
E. U. Condon
W. F. Fair, Jr.
Harvey Fletcher
George R. Harrison
W. V. Houston
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
John T. Tate
Wallace Waterfall
M. W. White
A. G. Worthing

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton, *Director*
Theodore Vorburger, *Business Manager*
Ruth F. Bryans, *Publication Manager*
Betty Hofstadter, *Office Manager*
Eileen B. Neuberger, *Circulation Manager*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $4 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices
Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 20, Number 5 May, 1949

The Triangulation of a Two-Dimensional Continuum for the Purpose of the Approximate Solution of Second-Order Partial Differential Equations

L. Tasny-Tschiassny
Department of Electrical Engineering, University of Sydney, N.S.W., Australia
(Received April 26, 1948)

It is shown how a non-homogeneous, non-isotropic, electrically conductive, two-dimensional continuum can be replaced by a network of resistors whereby the nodes of the network are situated at any arbitrarily selected points, and how Southwell's relaxation method can be applied to this network for the solution of potential problems.

I. THE GENERAL DIFFERENTIAL EQUATION CONSIDERED

SECOND-ORDER partial differential equations for the unknown function $V=V(x, y)$ of the general form

$$\frac{\partial}{\partial x}\left(\gamma_x\frac{\partial V}{\partial x}+\gamma_{xy}\frac{\partial V}{\partial y}\right)+\frac{\partial}{\partial y}\left(\gamma_{xy}\frac{\partial V}{\partial x}+\gamma_y\frac{\partial V}{\partial y}\right)-i(x, y)=0, \quad (1)$$

where γ_x, γ_{xy}, γ_y, and $i(x, y)$ are given functions of x and y, can be interpreted by the analog of a plane non-homogeneous and non-isotropic electrically conducting sheet, the conductivity of which (given by the quantities γ_x, γ_{xy}, γ_y) is a two-dimensional symmetrical tensor varying from point to point, and the transverse loading of which by an external electric current density leaving the conducting sheet is represented by the function $i(x, y)$ which varies from point to point. $V(x, y)$ is the electric potential and $-[\gamma_x(\partial V/\partial x)+\gamma_{xy}(\partial V/\partial y)]$ and $-[\gamma_{xy}(\partial V/\partial x)+\gamma_y(\partial V/\partial y)]$ are the components in the x and y directions of the current density.

Special cases of Eq. (1) are the quasi-plane-harmonic equation ($\gamma_x=\gamma_y$ and $\gamma_{xy}=0$), the two-dimensional Poisson equation ($\gamma_x=\gamma_y=$constant, $\gamma_{xy}=0$), and the two-dimensional Laplace Equation ($\gamma_x=\gamma_y=$constant, $\gamma_{xy}=0$, $i(x, y)=0$).

Methods of solving Eq. (1) approximately for given boundary conditions are very important for a large variety of mechanical and electrical problems.

II. THE REPLACEMENT OF A DIFFERENTIAL EQUATION BY A DIFFERENCE EQUATION

When solving a differential equation for the function $V(x, y)$ approximately as a difference equation, one concentrates on the values of the function $V(x, y)$ at a sufficiently large number of specific points n or s, called in the following the nodal points, and tries to find a set of values V_n or V_s for $V(x, y)$ at these points that complies

(*a*) with a difference equation, i.e., an equation that expresses a relation between the values of $V(x, y)$ at a certain nodal point and some neighboring nodal points,

(*b*) with the given boundary conditions.

The difference equation is derived from the differential equation by expressing the first derivative of a function as the quotient of the function values and the argument values at two neighboring points, and by applying this approximation to the higher derivatives in considering them as the first derivative of the derivative of one degree less.

Southwell's relaxation method is a modern and very efficient method for finding a set of values of $V(x, y)$ that complies with the difference equation.*

Southwell pointed out that Prandtl's Membrane** analogy when used as a physical representation of the quasi-harmonic, Poisson, and Laplace equations, i.e., special cases of Eq. (1), may be considered as being physically replaced by a network

* R. V. Southwell, *Relaxation Methods in Theoretical Physics* (Clarendon Press, Oxford, 1946).

** See reference 1, pp. 39, 86, 147.

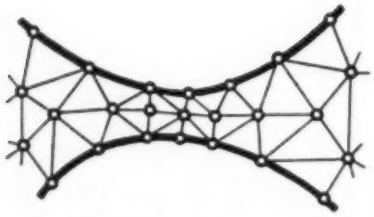

FIG. 1. Net of triangles of different shapes and sizes.

of strings connected at the nodal points, if the differential equation is replaced by the difference equation. In the conducting sheet analog used in this paper the conducting sheet may be considered as being physically replaced by a network of electric resistors joined at the nodal points. Expressed in terms of the conducting sheet analog the approximate solution of a differential equation by a difference equation may be described in the following way:

(1) Replace the two-dimensional continuum by a network of electric resistors. The value of the conductance G_{sn} of the resistor connecting the two nodes s and n depends on the type of the difference equation.

(2) Find suitable values of the transverse node loading currents I_s that are to be placed at the nodes in order to simulate the loading of the continuum by the continuous current density $i(x, y)$.

(3) Consider the potentials V_n and V_s of the nodes as unknowns, work out for each node s the expression

$$I_s = \sum_n (V_n - V_s) \cdot G_{ns}, \tag{2}$$

and put I_s equal to I_s for each node;

$$I_s = I_s. \tag{3}$$

The equations are simultaneous linear equations for the unknowns V_s. Since many of the coefficients G_{ns} are equal to nought, the equations are particularly suitable for solution by approximate methods.

As far as the author is aware, previous methods have used only arrangements of the nodal points in regular patterns of triangles, squares, and hexagons.*** Certain artifices were necessary to cope with the irregularities arising at the boundaries. Furthermore the general form of Eq. (1) in which $\gamma_x \neq \gamma_y$ and $\gamma_{xy} \neq 0$ has not been given special attention.

III. SOUTHWELL'S RELAXATION METHOD

Southwell's procedure for finding an approximate set of values V_s complying with Eqs. (2) and (3) is this:

(1) Approximate values of V_s are assumed for all nodes.

(2) I_s is computed for all nodes according to Eq. (2) and the values of the so-called residuals R_s are found for all nodes according to the formula

$$R_s = I_s - I_s. \tag{4}$$

(3) The potential of one node s with a large residual is changed by the value ΔV_s. The change of the residual ΔR_s of this node is found from the formula

$$\Delta R_s = -\Delta V_s \cdot \sum_n G_{sn}, \tag{5}$$

and of all other nodes n from the formula

$$\Delta R_n = \Delta V_s \cdot G_{ns}. \tag{6}$$

The values of the new residuals are computed.

(4) The procedure described in (3) is repeated for all nodes—always considering the new residuals—until all residuals have become negligible.

IV. THE TRIANGULATION OF THE CONTINUUM

The possibility of the replacement of the continuum by a network of regular pattern raises the question whether it would not be possible to employ for the same purpose a network of irregular triangles of different shapes and sizes with any specified position of the nodes (see Fig. 1). Once the values of the conductances of the resistors connecting the nodal points are known Southwell's relaxation method as described in Section III can be applied at once.

The advantages of an irregular pattern of nodes are: nodal points can be arranged for on the boundaries and on interlayers where the conductivity changes discontinuously (i.e., on the border lines of different materials), the nodes can be spaced differently at different parts of the field (grading of the net), and finally one can arrange for different directions of the lines connecting neighboring nodes at different parts of the field. The possibility of placing nodal points on the boundaries and interlayers eliminates all special problems connected with them, no matter whether the values of the potential or the values of the flux density are specified on the boundaries or whether the boundary conditions are mixed.

The author found that the question raised above can be answered affirmatively, and it is the purpose of the present paper to demonstrate this and to

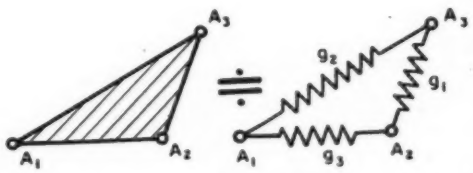

FIG. 2. Replacement of the interior of a triangle by a delta of resistors.

*** See, for example, the Historical Survey in reference 1, p. 36.

give the necessary mathematical formula for finding the values of the conductances between the nodes.

V. TRANSFORMATION FORMULAS FOR TENSOR CONDUCTIVITIES

Let (x, y) be the axes of an arbitrary system of Cartesian coordinates and (x', y') the axes of that system of coordinates that coincides at the point considered with the principal axes of the conductivity tensor at that point. Let ψ be the angle between the x' and x axes (see Fig. 3). Let $\gamma_{x'}$ and $\gamma_{y'}$ be the two different principal conductivities in the directions x' and y'. If $i_{x'}$, $i_{y'}$, i_x, i_y denote the components of the current density at the point considered and V the potential of this point, we obtain

$$-i_{x'}=\gamma_{x'}\cdot(\partial V/\partial x'), \tag{7}$$

$$-i_{y'}=\gamma_{y'}\cdot(\partial V/\partial y'). \tag{8}$$

It follows from the superposition principle and a transformation of the coordinates that

$$-i_x=\gamma_x\cdot\frac{\partial V}{\partial x}+\gamma_{xy}\cdot\frac{\partial V}{\partial y}, \tag{9}$$

$$-i_y=\gamma_{xy}\cdot\frac{\partial V}{\partial x}+\gamma_y\cdot\frac{\partial V}{\partial y}, \tag{10}$$

where

$$\gamma_x=\gamma_{x'}\cdot\cos^2\psi+\gamma_{y'}\cdot\sin^2\psi, \tag{11}$$

$$\gamma_y=\gamma_{x'}\cdot\sin^2\psi+\gamma_{y'}\cdot\cos^2\psi, \tag{12}$$

$$\gamma_{xy}=\frac{\gamma_{x'}-\gamma_{y'}}{2}\cdot\sin(2\psi). \tag{13}$$

By applying Kirchhoff's first rule to an infinitesimal element at the point considered we obtain Eq. (1) as the differential equation for the potential function V.

VI. THE LEADING PRINCIPLE IN THE DERIVATION OF THE BASIC FORMULA

In replacing Eq. (1) by a difference equation certain mathematical approximations are made. In the cases treated in the previous literature (i.e., for $\gamma_x=\gamma_y$, $\gamma_{xy}=0$ and for regular patterns formed by the nodal points) the replacement of the differential equation by a difference equation corresponds to the replacement of the continuum by a network of resistors (see Section II). The values of the resistors can be found either in the usual way from the difference equation, or, as can readily be demonstrated, from the concept of a physical network of resistors by the following procedure: It is assumed that the network is free from any transverse loading at the nodes and is placed in a uniform field. It is stipulated that for any intensity and direction of the field and for a certain width perpendicular to the direction of the field the average flux equals the actual flux produced in the continuum, if the continuum is not loaded by transverse current densities.

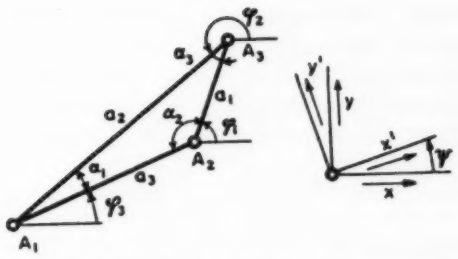

FIG. 3. Position of a triangle with respect to the axes (x, y) and (x', y').

The second procedure just described can be applied to the general case ($\gamma_x\neq\gamma_y$, $\gamma_{xy}\neq 0$, and for the nodes arranged in an irregular pattern) and is the method used in this paper for deriving the basic formula.

When carrying out the described procedure, one replaces the part of the continuum inside a triangle A_1, A_2, A_3 (Fig. 2) by a delta of resistors arranged between its vertices. If this is repeated for all triangles into which the continuum is divided, one always obtains two resistors in parallel that connect two nodal points, at least one of which is situated in the interior of the field, and one resistor between two nodal points situated on the boundary. By combining into one the two resistors connecting internal nodes, the final network that replaces the continuum is obtained.

VII. THE DERIVATION OF THE BASIC FORMULA

Let the triangle $A_1A_2A_3$ (Fig. 3) be subjected to a uniform field in an arbitrary direction. The components i_x and i_y of the current density in the plane are given by

$$i_x=-\left(\gamma_x\frac{\partial V}{\partial x}+\gamma_{xy}\frac{\partial V}{\partial y}\right), \tag{14}$$

$$i_y=-\left(\gamma_{xy}\frac{\partial V}{\partial x}+\gamma_y\frac{\partial V}{\partial y}\right). \tag{15}$$

The total current I_{12} leaving the interior of the triangle through the side A_1A_2 is given by

$$I_{12}=-a_3\sin\varphi_3\cdot\left(\gamma_x\frac{\partial V}{\partial x}+\gamma_{xy}\frac{\partial V}{\partial y}\right)+a_3\cos\varphi_3\cdot\left(\gamma_{xy}\frac{\partial V}{\partial x}+\gamma_y\frac{\partial V}{\partial y}\right), \tag{16}$$

and similar equations hold for the corresponding currents I_{23} and I_{31}.

If the continuum is to be equivalent to the delta of resistors, the current I_1 leaving the delta of resistors at the vertex A_1 must be given by

$$I_1=\tfrac{1}{2}(I_{12}+I_{31}), \tag{17}$$

and similar expressions hold for I_2 and I_3.

On the other hand, the symbols g_3, g_1, and g_2 denote the conductances of the resistors connected between the nodes A_1A_2, A_2A_3, and A_3A_1, respectively. The current I_1 is given by the expression

$$I_1=\left(a_3\cos\varphi_3\cdot\frac{\partial V}{\partial x}+a_3\sin\varphi_3\cdot\frac{\partial V}{\partial y}\right)\cdot g_3$$

$$-\left(a_2\cos\varphi_2\cdot\frac{\partial V}{\partial x}+a_2\sin\varphi_2\cdot\frac{\partial V}{\partial y}\right)\cdot g_2, \tag{18}$$

and the currents I_2 and I_3 by similar expressions. By substituting in (17) for I_{12} and I_{31} from the equations of the type of (16) and for I_1 from Eq. (17) we obtain an equation in which a simplification is possible by the help of

$$a_1\sin\varphi_1=-(a_2\sin\varphi_2+a_3\sin\varphi_3), \tag{19}$$

and

$$a_1\cos\varphi_1=-(a_2\cos\varphi_2+a_3\cos\varphi_3). \tag{20}$$

Since the resulting equation must hold for any values of $\partial V/\partial x$ and $\partial V/\partial y$, we obtain two equations by equating the coefficients of $\partial V/\partial x$ and $\partial V/\partial y$ on the right- and left-hand sides of the equation, respectively.

By eliminating g_2 and making consecutive use of the relations

$$\varphi_2-\varphi_3=\alpha_1+\pi, \tag{21}$$

$$\frac{a_1}{\sin\alpha_1}=\frac{a_3}{\sin\alpha_3}, \tag{22}$$

$$\varphi_1-\varphi_2=\alpha_3+\pi, \tag{23}$$

we obtain

$$2g_3=\frac{\gamma_x+\gamma_y}{2}\cdot\cot\alpha_3+\frac{\gamma_x-\gamma_y}{2}\cdot\frac{\cos(\varphi_1+\varphi_2)}{\sin\alpha_3}$$

$$+\gamma_{xy}\cdot\frac{\sin(\varphi_1+\varphi_2)}{\sin\alpha_3}. \tag{24}$$

By substituting for γ_x, γ_{xy}, and γ_y from Eqs. (11) to (13) we obtain as final result

$$2g_3=\frac{\gamma_{x'}+\gamma_{y'}}{2}\cdot\cot\alpha_3+\frac{\gamma_{x'}-\gamma_{y'}}{2}\cdot\frac{\cos(\varphi_1'+\varphi_2')}{\sin\alpha_3}, \tag{25}$$

where

$$\left.\begin{aligned}\varphi_1'&=\varphi_1-\psi,\\ \varphi_2'&=\varphi_2-\psi\end{aligned}\right\} \tag{26}$$

represent the inclinations of the sides towards the principal x' axis of the conductivity tensor. Hence g_3 is invariant. Similar equations hold for g_1 and g_2.

Mathematical investigations not described here show that equations of the type of (24) or (25) always result whatever starting point is taken for finding the values of g_1, g_2, and g_3 and that for an assembly of deltas of different sizes and shapes the resulting network replaces the continuum in every respect, as long as the field is uniform. Hence Eq. (25) is the basic formula in the sense of Section VI.

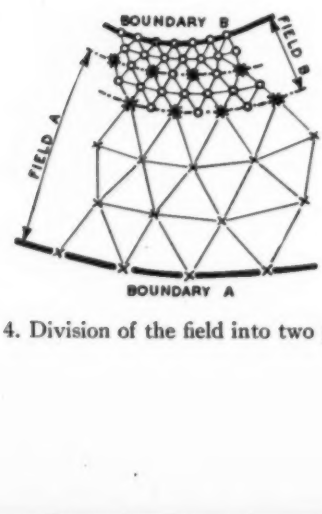

FIG. 4. Division of the field into two parts.

VIII. CONCLUSIONS FROM THE BASIC FORMULA

As was explained in Section VI, there are always two conductances to be considered that connect two nodes unless both nodes are on the boundary. These two conductances are given by Eq. (25) and will be called partial conductances. Their sum will be called total conductance between the two nodes concerned.

Formula (25) shows that the partial conductance consists of two additive terms, the first depending on the mean conductivity $(\gamma_{x'}+\gamma_{y'})/2=(\gamma_x+\gamma_y)/2$ and the second on the degree to which the material is non-isotropic, i.e., the value $(\gamma_{x'}-\gamma_{y'})/2$. Partial, and consequently total, conductances can be positive, zero, and negative. For isotropic materials $(\gamma_{x'}=\gamma_{y'})$ the partial conductance is positive, zero, and negative, if the angle opposite to the side considered of the triangle is acute, right, and obtuse, respectively.

Cyclic polygons in an isotropic medium exhibit interesting features. Let the polygon be divided into triangles drawn through one of its vertices. Two triangles adjacent to one diagonal form a cyclic quadrilateral. The two angles that are opposite to the common diagonal are supplementary, and therefore the total conductance corresponding to the

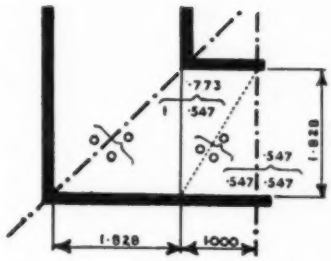

FIG. 5. Square-section cable in square sheath. No nodes between boundaries.

diagonal equals zero. Hence, a cyclic polygon in an isotropic medium is represented by a polygon of resistors along its sides without any resistor across the diagonals. This arrangement is known for squares and regular hexagons.

IX. NOTES FOR THE PRACTICAL USE OF THE METHOD WITH SPECIAL REGARD TO SOUTHWELL'S RELAXATION METHOD

The continuum between the boundaries is triangulated according to the viewpoints given in Section IV. If a boundary has sharp corners, nodes are placed at these corners. If the sharp corner is concave, the network is arranged in such a way that the total conductance of a resistor connecting the corner node with a node inside the medium is nearly zero. At convex sharp corners the conductance of resistors leaving the corner between the two perpendiculars to the boundary should not differ very much from zero.

As to the conductivity, a reference value equal to unity may be assumed and an appropriate multiplying factor be applied in the final result. For nonhomogeneous media the value of the conductivities at the center of gravity of each triangle considered or the arithmetic mean of the values of the conductivities at its vertices is substituted in Eq. (25). For computing partial conductances or for checking their values the relation

$$\cot\alpha_3=\frac{1-\cot\alpha_1\cdot\cot\alpha_2}{\cot\alpha_1+\cot\alpha_2} \qquad (27)$$

is useful, where α_1, α_2, and α_3 are the angles of a triangle. Problems in connection with three-dimensional fields of axial symmetry can be solved by considering the plane field of a section containing the axis and substituting in (25) for $\gamma_{x'}$ and $\gamma_{y'}$ the product of the actual conductivities at the individual points and the distance of the points from the axis. If the conductivity at a point depends on the current density at that point, as is the case in the analog of magnetic flux problems in ferromagnetic materials, successive approximations lead to the final solution.

Regarding transverse loading by current densities two simple procedures are possible. One is to place at each vertex of a triangle a load equal to one third of the area of the triangle multiplied by the current density of the loading at its center of gravity and to sum up the loads at each node. The other is to sum up the areas of all triangles meeting at a certain node and to multiply one third of the total of the areas by the current density of the loading at the node. Cyclic or regular polygons are to be divided into triangles for this purpose.

If Southwell's relaxation method is used, one can work on a graph of the net, as is customary in this method. The values of the total conductances between the nodes and the sum of the total conductances meeting at a node are put down once and for all, generally in ink. The potentials of the nodes, the currents established in the conductances between the nodes, and the residuals at the nodes are put down in pencil and amended during the process of the liquidation of residuals.

As to the grading of the net, a special procedure which is a generalization of an artifice developed by Southwell† should be given and explained in terms of the conducting sheet analog. The continuum is divided into two parts by two sets of arbitrarily selected reference nodes which lie on two contours approximately parallel and reasonably close to each other. Figure 4 shows an example of the double set of nodes. They are marked by asterisks. The procedure is to split up the field problem into two separate field problems, one being the problem of field A between the boundary A and the set of reference nodes nearer to the boundary B; and the other being the problem of Field B between the boundary B and the set of reference nodes nearer to the boundary A. The nodes for field A are the nodes marked by crosses and asterisks and the nodes for field B are the nodes marked by circles and asterisks. The two problems are solved separately, the only link being that the potentials of the nodes

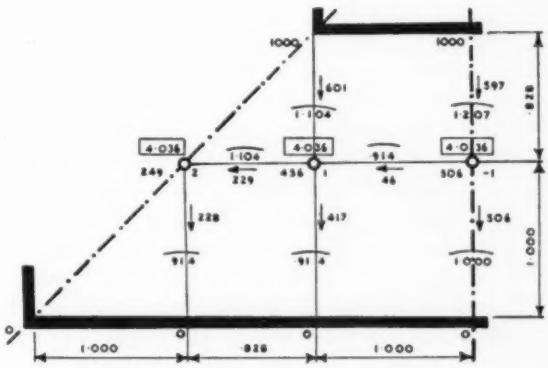

FIG. 6. Square-section cable in square sheath. Three nodes between boundaries.

† See reference 1, p. 98 and Fig. 40.

marked by asterisks must be the same in both partial problems. By this arrangement the requirement is met that the values of the potentials and of the gradients in the two partial problems are approximately equal near the reference nodes. The solutions of the two partial problems taken together are the solution of the main problem.

X. EXAMPLE

As an example, the capacity of a square-section cable in a square sheath with an homogeneous isotropic dielectric of dielectric constant unity will be computed. This example was treated by Gandy and Southwell.†† The ratio of the two sides of the square is $2\cdot\sqrt{2}$. Figure 5 shows the triangulation of one-eighth of the cross section without any nodes between the boundaries. In the graph double the values of the partial conductances and the arithmetic means of these values, i.e., the total conductances are entered. If C is the required capacity, one obtains at once without any liquidation of residuals $4\pi C=4\times0.547+8\times0.773=8.37_2$. The value given by Gandy and Southwell is $4\pi C=6.71_5$, or the error is +24.7 percent.

Figure 6 shows the treatment of the same problem with three nodes between the boundaries, and the final solution found by the relaxation process is entered in the figure. The numbers under arcs are total conductances, the numbers in rectangles their sums, the numbers at the left-hand sides of nodes the potentials of the nodes, the numbers near arrows the currents established in the total conductances, and the numbers at the right-hand sides of nodes the residual currents. The capacity of the cable can be found either from the total current leaving the conductor or the total current entering the sheath. The first alternative leads to

$$4\pi C=\frac{8\times601+4\times597}{1000}=7.196,$$

and the second to

$$4\pi C=\frac{8\times(228+417)+4\times506}{1000}=7.184.$$

The deviation of the mean of $4\pi C=7.19_0$ from Gandy's and Southwell's value is +7.1 percent.

The author computed the capacity for different arrangements and different numbers of nodes. For 6 nodes and a triangulation that gave closer meshes near the vertex of the inner square than at other points of the field the error was +3.5 percent. For a square net without grading and 15 nodes the error was +2 percent.

XI. CONCLUSION

A procedure has been developed according to which any differential equation of the type of Eq. (1) can be solved approximately by a relaxation method whereby the use of only five different formulas is necessary, *viz.*, the use of formulas (2), (4)–(6), and (24) or (25), as the case may be. The shape of the boundary and the type of the boundary conditions are irrelevant as far as the procedure is concerned.

†† See reference 1, p. 133, and R. W. C. Gandy and R. V. Southwell, Phil. Trans. Roy. Soc. **A238**, 453–75 (1940).

On the Velocity-Dependent Characteristics of High Frequency Tubes

JULIAN K. KNIPP
Department of Physics and Institute for Atomic Research, Iowa State College, Ames, Iowa
(Received July 9, 1948)

The electronic transadmittances, transmodulation functions, and other characteristics of a planar, multigrid high frequency tube are expressed in terms of four fundamental admittance and modulation functions which are sufficient for the description of the low level r-f properties of the electron beam. A matrix representation is introduced based on the successive velocity moments of the beam which leads to the formulation of the characteristics as products involving excitation, grid transmission, modulation, and modulation admittance matrices. The electronic equations obtained are a formulation in which the Llewellyn approximation that the second and all higher velocity moments are zero is not made. Thus the theory allows for higher order velocity effects but is dependent on the knowledge of the fundamental admittance and modulation functions.

I. INTRODUCTION

THE pioneer work of Benham[1] and Llewellyn[2] on transit effects in electron beams opened the field of velocity-dependent characteristics, which characteristics are associated with all electronic high frequency phenomena. Velocity modulation soon became a valuable technique at all power levels of operation.[3] In the low level field (small modulation amplitudes) Llewellyn made the remarkable achievement of combining density and velocity modulation effects in a planar diode in a manner completely consistent with electron dynamics and field divergence.[4] This is a single-velocity theory which does not include effects of velocity spread.

The electronic characteristics of a planar grid tube can be expressed in terms of certain fundamental admittance and modulation functions $y_m(x,v)$, $y_m{}^1(x)$, $\mu(x,v;v')$, $\mu^1(x;v')$, which are sufficient for the description of the r-f components of electron beam within a single region of the tube. Under conditions of low level operation, which are assumed in this paper, the r-f current density at a specified plane parallel to the grids resulting from electrons of velocity v is the sum of a current density proportional to the r-f voltage of the region and a current density which arises from modulations present in the beam as it enters the region. Thus, within the region a distance x from its first bounding plane, the r-f current density for electrons of velocity v is

$$i_\omega(x,v) = -y_m(x,v)V_\omega + i_m(x,v), \qquad (1)$$

where V_ω is the r-f gap voltage, $y_m(x,v)$ is the *modulation admittance function* at this plane for electrons of this velocity, and $i_m(x,v)$ is the *modulation current density* just mentioned.[5] If electrons enter only at the plane $x=0$, a condition which is also assumed throughout this paper, the modulation current density is an expression of the form

$$i_m(x,v) = \int_{v^1(0)}^{\infty} dv'\mu(x,v;v')i_\omega(0,v') - \mu(x,v;v^1(0))i(0,v^1(0))v_\omega{}^1(0), \qquad (2)$$

where $\mu(x,v;v')$ is the *density modulation function* at the plane x for electrons of velocity v arising from modulations due to electrons injected with velocity v'; the d.c. current density at the injection plane is $i(0,v')$, and the quantities $v^1(0)$ and $v_\omega{}^1(0)$ are the d.c. and r-f components, respectively, of the lowest velocity with which electrons enter the region (if the injection plane is the cathode, $v^1(0)$ and $v_\omega{}^1(0)$ are zero).

The first problem of a complete theory is the calculation of the admittance and modulation functions from the dynamics of the electrons and their contributions to the fields. Since the charge density at any plane depends on the electrons arriving at that plane, and hence on the electric fields through which they have passed since injection, the electric field appears on both sides of the divergence equation relating charge density and field. This equation is, in fact, an integral equation for the field, which in the small signal approximation breaks up into an integral equation for the determination of the d.c. field and an integral equation linear in r-f amplitudes for the determination of the r-f field. The admittance and modulation functions follow directly from a knowledge of these fields. It is assumed that this problem has been solved and that these functions are known.[6]

In the absence of the r-f fields, the motion of the electrons is controlled by the emission characteris-

[1] W. E. Benham, Phil. Mag. **5**, 641 (1928); **11**, 457 (1934).
[2] F. B. Llewellyn, Bell Sys. Tech. J. **14**, 632 (1935); **15**, 575 (1936).
[3] W. C. Hahn, G. E. Review **42**, 258 (1939); W. C. Hahn and G. E. Metcalf, Proc. I.R.E. **27**, 106 (1939). R. H. Varian and S. F. Varian, J. App. Phys. **10**, 321 (1939). D. L. Webster, J. App. Phys. **10**, 501 (1939); **10**, 864 (1939).
[4] F. B. Llewellyn, *Electron Inertia Effects* (Cambridge University Press, London, 1939).
[5] See chapters 3 and 5 of D. R. Hamilton, J. K. Knipp, and J. B. H. Kuper, *Klystrons and Microwave Triodes* (McGraw-Hill Book Company, Inc., New York, 1948).
[6] For a discussion, see the last section of this paper.

tics of the cathode and the d.c. voltages on the grids and anode. In order to insure uniformity over planes parallel to the grids, a uniformly emitting cathode, shallow regions, and reasonably fine mesh grids are assumed. Because of d.c. field penetration of the grid, it is possible, even if the control grid is negative, that as far as the motion of the electrons is concerned the effective voltage of that grid is positive; but nevertheless, since it is negative, electron collection by that grid is practically zero. It is assumed, however, that there is no r-f penetration of the grids. The action of grid $(k-1)$ on the r-f current density is described by

$$i_\omega(0,v)^{(k)}=\alpha(v)^{(k-1)}i_\omega(d,v)^{(k-1)}+i_g(v)^{(k-1)}, \quad (3)$$

where $i_\omega(0,v)^{(k)}$ is the injected r-f current density in the region (k), $i_\omega(d,v)^{(k-1)}$ is the r-f current density at the exit plane of region $(k-1)$, $\alpha(v)^{(k-1)}$ is the velocity-dependent *transmission function* of grid $(k-1)$, separating the two regions, and $i_g(v)^{(k-1)}$ is the *grid modulation current density* introduced into the beam by that grid. The grid characteristics $\alpha(v)$ and $i_g(v)$ are determined by the geometry and nature of the fields in the neighborhood of the grid and by the details of the process of capture and the production of secondaries.

II. BEAM AND MODULATION CURRENTS

The r-f electron current at the plane x within a particular region has the form

$$i_\omega(x)=-y_m(x)V_\omega+i_m(x), \quad (4)$$

where $y_m(x)$ is the density *modulation admittance* at the plane x and $i_m(x)$ is the *modulation current* at that plane arising from modulations present in the beam at injection. These quantities can be expressed in terms of the admittance and modulation functions.

It is first to be noted that Eq. (1) must be supplemented by an expression for the r-f component of the lowest (or most negative) velocity at that plane. The required relation can be written in the form

$$v_\omega{}^l(x)=-y_m{}^l(x)\left(\frac{v^l(x)}{-I_0}\right)V_\omega+v_m{}^l(x), \quad (1a)$$

in which $v^l(x)$ is the d.c. component of this velocity, and $-I_0$ is the d.c. beam current in the region; the first term is the modulation in the lowest velocity proportional to the gap voltage and the second term is the modulation arising from modulations present in the beam at injection. The latter has the form

$$v_m{}^l(x)=\frac{v^l(x)}{-I_0}\left\{\int_{v^l(0)}^{\infty}dv'\mu^l(x;v')i_\omega(0,v') -\mu^l(x;v^l(0))i(0,v^l(0))v_\omega{}^l(0)\right\}. \quad (2a)$$

Equations (1) and (1a), together with Eqs. (2) and (2a), give a complete description of the r-f properties of the beam at the plane x in terms of the gap voltage V_ω and $i_\omega(0,v')$ and $v_\omega{}^l(0)$, which describe the r-f properties at the injection plane.

Substitution of Eqs. (1) and (1a) in

$$i_\omega(x)=\int_{v^l(x)}^{\infty}dvi_\omega(x,v)-i(x,v^l(x))v_\omega{}^l(x), \quad (5)$$

and comparison with Eq. (4) yields for the density modulation admittance and the modulation current, respectively,

$$y_m(x)=\int_{v^l(x)}^{\infty}dvy_m(x,v) -i(x,v^l(x))y_m{}^l(x)(v^l(x)/-I_0),$$

$$i_m(x)=\int_{v^l(x)}^{\infty}dvi_m(x,v)-i(x,v^l(x))v_m{}^l(x). \quad (6)$$

Substitution of Eqs. (2) and (2a) in the second of Eqs. (6) gives

$$i_m(x)=\int_{v^l(0)}^{\infty}dv'\mu(x;v')i_\omega(0,v') -\mu(x;v^l(0))i(0,v^l(0))v_\omega{}^l(0), \quad (7)$$

with the *modulation function*,

$$\mu(x;v')=\int_{v^l(x)}^{\infty}dv\mu(x,v;v') -i(x,v^l(x))(v^l(x)/-I_0)\mu^l(x;v'). \quad (8)$$

III. DRIVING AND EXCITING CURRENTS

The circuits of a high frequency tube are usually cavity resonators with the gaps between the cathode, grids, and anode forming the capacitances. Since the regions are shallow, the fields in a particular gap in the absence of the beam are nearly space-constant and the beam is coupled to the resonator and load through the space average of the electron current in the gap. This average is called the *driving current* i_ω; thus

$$i_\omega=1/d\int_0^d dxi_\omega(x). \quad (9)$$

If Y_R and Y_L are the resonator and load admittances, respectively, measured at the gap, $i_\omega=(Y_R+Y_L)V_\omega$. The space average of Eq. (4) gives

$$i_\omega=-Y_BV_\omega+i_m, \quad (10)$$

where

$$Y_B=1/d\int_0^d dxy_m(x), \quad i_m=1/d\int_0^d dxi_m(x), \quad (11)$$

are the *beam-loading admittance* and *exciting current*, respectively. If the total circuit admittance is $Y=Y_B+Y_R+Y_L$, the gap voltage arising from modulations in the beam at injection is given by the simple relation $i_m=YV_\omega$. The exciting current is given by

$$i_m=\int_{v^l(0)}^{\infty}dv'\mu(v')i_\omega(0,v') -\mu(v^l(0))i(0,v^l(0))v_\omega{}^l(0), \quad (12)$$

where $\mu(v')$ is the *excitation function* for modulations in the injected beam due to electrons having velocities v':

$$\mu(v')=1/d\int_0^d dx\mu(x;v'). \quad (13)$$

IV. EXCITING CURRENT IN THE CATHODE REGION

Since the electrons emitted by the cathode have all velocities from 0 to ∞, the exciting current in the cathode region is simply

$$i_m{}^{(c)}=\int_0^{\infty}dv'\mu(v')^{(c)}i_\omega(0,v')^{(c)}. \quad (14)$$

The cathode excitation function $\mu(v')^{(c)}$ is the space average of the current modulation function $\mu(x;v')^{(c)}$, which as in Eq. (8) can be expressed in terms of the modulation functions $\mu(x,v;v')^{(c)}$ and $\mu^l(x;v')^{(c)}$, defined by Eqs. (2) and (2a). In the cathode region electrons may be reflected from a potential minimum as a result of space charge or from a retarding field arising from a negative effective voltage on the control grid. Hence at a given plane in the region, the velocity distribution may extend down to a lower negative value; such is implied in Eq. (8). The definitions of Eqs. (2) and (2a) allow for this possibility.

The beam-loading admittance of the cathode region $Y_B{}^{(c)}$ is the space average of $y_m(x)^{(c)}$ as given by the first of Eqs. (6). The modulation admittances $y_m(x,v)$ and $y_m{}^l(x)$ are used without modification as defined in Eqs. (1) and (1a).

V. TRANSFER RELATIONS FROM REGIONS $(k-1)$ TO (k)

Separating regions (k) and $(k-1)$ is grid $(k-1)$. The action of the grid is described in the Introduction, to which description is added the assumption that the r-f components of the velocity of the slowest electrons approaching the grid and the slowest electrons leaving it after transmission are identical. This limit does not apply to the grid modulation current density $i_g(v)^{(k-1)}$, which has the lower limit 0. Hence

$$i_\omega(0,v')^{(k)}=\alpha(v')^{(k-1)}i_\omega(d,v')^{(k-1)}+i_g(v')^{(k-1)},$$
$$v_\omega{}^l(0)^{(k)}=v_\omega{}^l(d)^{(k-1)}. \quad (15)$$

These relations can be used in Eq. (12) for the exciting current,

$$i_m{}^{(k)}=\int_{v^l(0)^{(k)}}^{\infty}dv'\mu(v')^{(k)}\alpha(v')^{(k-1)}i_\omega(d,v')^{(k-1)} -\mu(v^l(0))^{(k)}i(0,v^l(0))^{(k)}v_\omega{}^l(d)^{(k-1)} +\int_0^{\infty}dv'\mu(v')^{(k)}i_g(v')^{(k-1)}. \quad (16)$$

On using Eqs. (1) and (1a) for region $(k-1)$, the exciting current of region (k) is transformed into the sum of a current proportional to the gap voltage of region $(k-1)$, an exciting current dependent on the exciting current of region $(k-1)$, and an exciting current arising from grid $(k-1)$. The result has the form

$$i_m{}^{(k)}=-Y_m{}^{(k,\,k-1)}V_\omega{}^{(k-1)}+i_m{}^{(k,\,k-1)}+i_g{}^{(k,\,k-1)}, \quad (17)$$

in which

$$Y_m{}^{(k,\,k-1)}=\int_{v^l(0)}^{\infty}dv'\mu(v')^{(k)}\alpha(v')^{(k-1)}y_m(d,v')^{(k-1)} -\mu(v^l(0))^{(k)}i(0,v^l(0))^{(k)}y_m{}^l(d)^{(k-1)} \times(v^l(d)/-I_0)^{(k-1)},$$

$$i_m{}^{(k,\,k-1)}=\int_{v^l(0)}^{\infty}dv'\mu(v')^{(k)}\alpha(v')^{(k-1)}i_m(d,v')^{(k-1)} -\mu(v^l(0))^{(k)}i(0,v^l(0))^{(k)}v_m{}^l(d)^{(k-1)}, \quad (18)$$

$$i_g{}^{(k,\,k-1)}=\int_0^{\infty}dv'\mu(v')^{(k)}i_g(v')^{(k-1)},$$

are, respectively, the electronic *transadmittance* from region $(k-1)$ to region (k), the *transexciting current* from region $(k-1)$ in region (k), and the *grid exciting current* from grid $(k-1)$ in region (k).

The use of Eqs. (2) and (2′) yields

$$i_m{}^{(k,\,k-1)}=\int_{v^l(0)^{(k-1)}}^{\infty}dv''\mu(v'')^{(k,\,k-1)}i_\omega(0,v'')^{(k-1)} -\mu(v^l(0))^{(k,\,k-1)}i(0,v^l(0))^{(k-1)}v_\omega{}^l(0)^{(k-1)}, \quad (19)$$

in which

$$\mu(v'')^{(k,\,k-1)}=\int_{v^l(0)^{(k)}}^{\infty}dv'\mu(v')^{(k)}\alpha(v')^{(k)} \times\mu(d,v';v'')^{(k-1)}-\mu(v^l(0))^{(k)}i(0,v^l(0))^{(k)} \times(v^l(d)/-I_0)^{(k-1)}\mu^l(d;v')^{(k-1)} \quad (20)$$

is the *transmodulation function* from region $(k-1)$ to region (k).

VI. BEAM CURRENT WITHOUT REFLECTIONS

If the beam at a given plane contains no electrons with negative velocities, it is convenient to describe it by giving the beam current, instantaneous average velocity, and instantaneous velocity moments. Thus if $i(x,v,t)$ is the current density and if $v^l(x,t)$ is the lower limit of the velocity distribution, the *beam current* $i(x,t)$ and instantaneous *average velocity* $v(x,t)$ are given by

$$i(x,t)=\int_{v^l(x,t)}^{\infty} dvi(x,v,t),$$

$$v(x,t)i(x,t)=\int_{v^l(x,t)}^{\infty} dvvi(x,v,t), \tag{21}$$

and the nth instantaneous *velocity moment* $\Delta^{(n)}(x,t)$ by

$$\Delta^{(n)}(x,t)i(x,t)=\int_{v^l(x,t)}^{\infty} dv[v(x,t)-v]^n i(x,v,t). \tag{22}$$

One notes that the zeroth moment is unity and the first moment is zero.

For the purpose of calculating average values the current density can be represented by

$$i(x,v,t)=i(x,t)\delta(x,v,t), \tag{23}$$

where

$$\delta(x,v,t)=\sum_0^\infty \frac{1}{n!}\Delta^{(n)}(x,t)\delta^{(n)}(v-v(x,t)). \tag{24}$$

In this expression $\delta^{(n)}(v-v(x,t))$ is the nth derivative of the Dirac delta-function. On substituting this expression in Eq. (22) it is seen at once that all the moments are given correctly.

It follows that the representation of the d.c. and r-f components of the current density are, respectively,

$$i(x,v)=-I_0\delta(x,v),$$

$$i_\omega(x,v)=i_\omega(x)\delta(x,v)+I_0v_\omega(x)\delta'(x,v)$$

$$-I_0\sum_2^\infty \frac{1}{n!}\Delta_\omega^{(n)}(x)\delta^{(n)}(x-v(x)), \tag{25}$$

where

$$\delta(x,v)=\sum_0^\infty \frac{1}{n!}\Delta^{(n)}(x)\delta^{(n)}(v-v(x)), \tag{26}$$

and $\delta'(x,v)$ is the first velocity derivative of $\delta(x,v)$. In these expressions $\Delta^{(n)}(x)$ and $\Delta_\omega^{(n)}(x)$ are the d.c. and r-f components of the nth velocity moment, respectively. Since Eqs. (25) contain the velocity only through the delta-function and its derivatives, modulations of the lower velocity limit are to be neglected in using this representation for the d.c. and r-f current densities.

At this point it proves to be very convenient to introduce a matrix representation for $i_\omega(x,v)$ which effects a separation of velocity dependence and r-f moments. This is done by writing $i_\omega(x,v)$ as the product of two matrices, thus,

$$i_\omega(x,v)=\boldsymbol{\delta}(x,v)\mathbf{i}_\omega(x), \tag{27}$$

where

$$\boldsymbol{\delta}(x,v)=|\,\delta(x,v),\ -v(x)\delta'(x,v),\ \tfrac{1}{2}\Delta^{(2)}(x)\delta^{(2)}(v-v(x)),\ \tfrac{1}{6}\Delta^{(3)}(x)\delta^{(3)}(v-v(x)),\ \cdots\,| \tag{28}$$

is a dimensionless one-row matrix, and

$$\mathbf{i}_\omega(x)=-I_0\begin{vmatrix} i_\omega(x)/-I_0 \\ v_\omega(x)/v(x) \\ \Delta_\omega^{(2)}(x)/\Delta^{(2)}(x) \\ \Delta_\omega^{(3)}(x)/\Delta^{(3)}(x) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} \tag{29}$$

is a one-column *current operand.*

The current operand can be written in the form

$$\mathbf{i}_\omega(x)=-\mathbf{y}_m(x)V_\omega+\mathbf{i}_m(x), \tag{30}$$

where $\mathbf{y}_m(x)$ is the modulation *admittance matrix* and $\mathbf{i}_m(x)$ is the *modulation current operand* arising from the modulations present in the beam at injection. These two matrices are readily expressed in terms of $y_m(x,v)$, $i_m(x,v)$, $y_m{}^l(x)$, $v_m{}^l(x)$ of Eqs. (1) and (1a). In fact

$$\mathbf{y}_m(x)=\int_{v^l(x)}^{\infty} dvy_m(x,v)\mathbf{1}(x,v)$$

$$-i(x,v^l(x))y_m{}^l(x)(v^l(x)/-I_0)\mathbf{1}(x,v^l(x)),$$

$$\mathbf{i}_m(x)=\int_{v^l(x)}^{\infty} dvi_m(x,v)\mathbf{1}(x,v)$$

$$-i(x,v^l(x))v_m{}^l(x)\mathbf{1}(x,v^l(x)), \tag{31}$$

where

$$\mathbf{1}(x,v)=\begin{vmatrix} 1 \\ -(v(x)-v)/v(x) \\ [(v(x)-v)^2-\Delta^{(2)}(x)]/\Delta^{(2)}(x) \\ [(v(x)-v)^3-\Delta^{(3)}(x) \\ \quad -3\Delta^{(2)}(x)(v(x)-v)]/\Delta^{(3)}(x) \\ [(v(x)-v)^4-\Delta^{(4)}(x) \\ \quad -4\Delta^{(3)}(x)(v(x)-v)]/\Delta^{(4)}(x) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{vmatrix} \tag{32}$$

is a dimensionless one-column matrix. It has the

following properties:

$$\mathfrak{d}(x,v)\mathfrak{1}(x,v')=\delta(v-v'), \qquad \int dv\mathfrak{1}(x,v)\mathfrak{d}(x,v)=1, \quad (33)$$

where $\mathfrak{1}$ is the identity matrix.

The modulation current operand can be written in terms of the properties of the beam at the entrance plane by using Eqs. (2) and (2a). Thus

$$\mathfrak{i}_m(x)=\int_{v^l(0)}^{\infty} dv'\mathfrak{y}(x;v')i_\omega(0,v') \\ -i(0,v^l(0))v_\omega{}^l(0)\mathfrak{y}(x;v^l(0)), \quad (34)$$

where

$$\mathfrak{y}(x;v')=\int_{v^l(x)}^{\infty} dv\mu(x,v;v')\mathfrak{1}(x,v) \\ -i(x,v^l(x))\left(\frac{v^l(x)}{-I_0}\right)\mu^l(x;v')\mathfrak{1}(x,v^l(x)), \quad (35)$$

is the *modulation matrix function.*

VII. MODULATION AND EXCITATION MATRICES

It was assumed in the last section that the beam at the plane x contained no electrons with negative velocities. The r-f current density was then represented as the product of two matrices, Eq. (27). If there are no negative-velocity electrons at the entrance plane, the r-f current density at that plane can be given the same matrix representation. When introduced into Eq. (33), there results

$$\mathfrak{i}_m(x)=\mathfrak{y}(x)\mathfrak{i}_\omega(0), \quad (36)$$

where

$$\mathfrak{y}(x)=\int dv'\mathfrak{y}(x;v')\mathfrak{d}(0,v') \quad (37)$$

is the doubly infinite (square) *modulation matrix.*

Similarly, from Eq. (12) for the exciting current,

$$i_m=\mathfrak{y}\mathfrak{i}_\omega(0), \quad (38)$$

where

$$\mathfrak{y}=\int dv'\mu(v')\mathfrak{d}(0,v') \quad (39)$$

is the one-row *excitation matrix.*

VIII. TRANSMISSION AND GRID-CURRENT MATRICES

If region (k) has no electrons with negative velocities, $i_\omega(0)^{(k)}=i_m(0)^{(k)}$. Hence on combining Eqs. (34) and (15) and introducing the matrix representation of $i_\omega(d,v')^{(k-1)}$ and observing that $\mathfrak{y}(0;v')=\mathfrak{1}(0,v')$, there results

$$\mathfrak{i}_\omega(0)^{(k)}=\mathfrak{a}^{(k-1)}\mathfrak{i}_\omega(d)^{(k-1)}+\mathfrak{i}_g{}^{(k-1)}, \quad (40)$$

in which

$$\mathfrak{a}^{(k-1)}=\int dv'\mathfrak{1}(0,v')^{(k)}\alpha(v')^{(k-1)}\mathfrak{d}(d,v')^{(k-1)} \quad (41)$$

is the *transmission matrix* for grid $(k-1)$ and

$$\mathfrak{i}_g{}^{(k-1)}=\int_0^{\infty} dv'\mathfrak{1}(0,v')^{(k)}i_g(v')^{(k-1)} \quad (42)$$

is the *grid-current operand.*

If the transmission function $\alpha(v')^{(k-1)}$ is a constant, $\mathfrak{1}(0,v')^{(k)}=\mathfrak{1}(d,v')^{(k-1)}$, and the transmission matrix is a constant, by the second of Eqs. (33).

IX. TRANSFER RELATIONS FROM REGIONS $(k-1)$ TO (k)

By Eq. (38), the exciting current for region (k) is $i_m{}^{(k)}=\mathfrak{y}^{(k)}\mathfrak{i}_\omega(0)^{(k)}$. With the aid of Eq. (40), $i_m{}^{(k)}=\mathfrak{y}^{(k)}[\mathfrak{a}^{(k-1)}\mathfrak{i}_\omega(d)^{(k-1)}+\mathfrak{i}_g{}^{(k-1)}]$. Hence, by using Eq. (30),

$$i_m{}^{(k)}=-\mathfrak{y}^{(k)}\mathfrak{a}^{(k-1)}\mathfrak{y}_m(d)^{(k-1)}V_\omega{}^{(k-1)} \\ +\mathfrak{y}^{(k)}\mathfrak{a}^{(k-1)}\mathfrak{i}_m(d)^{(k-1)}+\mathfrak{y}^{(k)}\mathfrak{i}_g{}^{(k-1)}. \quad (43)$$

This matrix relation is of the same form as Eq. (17). The electronic transadmittance, transexciting current, and grid exciting current are, respectively,

$$Y_m{}^{(k,k-1)}=\mathfrak{y}^{(k)}\mathfrak{a}^{(k-1)}\mathfrak{y}_m(d)^{(k-1)}, \\ i_m{}^{(k,k-1)}=\mathfrak{y}^{(k)}\mathfrak{a}^{(k-1)}\mathfrak{i}_m(d)^{(k-1)}, \\ i_g{}^{(k,k-1)}=\mathfrak{y}^{(k)}\mathfrak{i}_g{}^{(k-1)}. \quad (44)$$

That Eqs. (44) are completely equivalent to Eqs. (18) can be shown by the substitution into the former Eqs. (39), (41), (31), (39), (42) and the use of the first of Eqs. (33). Since $i(0,v^l(0))^{(k)}=\alpha(v^l(0))^{(k-1)}i(0,v^l(0))^{(k-1)}$, Eqs. (18) follow as written.

If region $(k-1)$ has no negative-velocity electrons, the use of Eq. (36) gives

$$i_m{}^{(k,k-1)}=\mathfrak{y}_m{}^{(k,k-1)}\mathfrak{i}_\omega(0)^{(k-1)}, \quad (45)$$

where

$$\mathfrak{y}_m{}^{(k,k-1)}=\mathfrak{y}^{(k)}\mathfrak{a}^{(k-1)}\mathfrak{y}(d)^{(k-1)} \quad (46)$$

is the *transmodulation matrix* from region $(k-1)$ to region (k). Eq. (45) also follows from Eqs. (19) and (20).

In the cathode region it may prove necessary to use the more general Eq. (34), which yields for the transexciting current to the second region

$$i_m{}^{(2,c)}=\int_0^{\infty} dv'\mu_m(v')^{(2,c)}i_\omega(0,v')^{(c)}, \quad (47)$$

where the transmodulation function is expressed by

$$\mu_m(v')^{(2,c)}=\mathfrak{y}^{(2)}\mathfrak{a}^{(1)}\mathfrak{y}(d;v')^{(c)}. \quad (48)$$

X. TRANSMODULATION MATRICES, TRANSADMITTANCES, TRANSMODULATION FUNCTIONS

Repeated use of the relations of the last three sections leads to the following general equation for the r-f exciting current in region (k):

$$\mathbf{i}_m^{(k)} = \sum_{j=i}^{k-1} \left[-Y_m^{(k,j)} V_\omega^{(j)} + \mathbf{\mu}_m^{(k,j)} \mathbf{i}_g^{(j)} \right] + \mathbf{\mu}_m^{(k,i)} \mathbf{\alpha}^{(i)} \mathbf{i}_m(d)^{(i)}. \quad (49)$$

In this expression $\mathbf{\mu}_m^{(k,j)}$ is the transmodulation matrix from grid (j) to region (k). It is given by

$$\mathbf{\mu}_m^{(k,j)} = \mathbf{\mu}^{(k)} \mathbf{\alpha}^{(k-1)} \mathbf{\mu}(d)^{(k-1)} \mathbf{\alpha}^{(k-2)} \mathbf{\mu}(d)^{(k-2)} \cdots \mathbf{\alpha}^{(j+1)} \mathbf{\mu}(d)^{(j+1)}. \quad (50)$$

The transadmittance $Y_m^{(k,j)}$ from region (j) to region (k) is simply

$$Y_m^{(k,j)} = \mathbf{\mu}_m^{(k,j)} \mathbf{\alpha}^{(j)} \mathbf{y}_m(d)^{(j)}. \quad (51)$$

In the last term of Eq. (49), $\mathbf{i}_m(d)^{(i)}$ can be substituted as calculated from Eq. (34). The term can then be written in the form

$$\int_{v^l(0)^{(i)}}^{\infty} dv' \mu_m(v')^{(k,i)} i_\omega(0,v')^{(i)} - i(0,v^l(0))^{(i)} v_\omega{}^l(0)^{(i)} \mu_m(v^l(0))^{(k,i)}, \quad (52)$$

where $\mu_m(v')^{(k,i)}$ is the transmodulation function from the entrance plane of region (i) to region (k). It is

$$\mu_m(v')^{(k,i)} = \mathbf{\mu}_m^{(k,i)} \mathbf{\alpha}^{(i)} \mathbf{\mu}(d;v')^{(i)}. \quad (53)$$

The fully detailed exciting current of region (k) is

$$\mathbf{i}_m^{(k)} = \sum_{j=1}^{k-1} \left[-Y_m^{(k,j)} V_\omega^{(j)} + \mathbf{\mu}_m^{(k,j)} \mathbf{i}_g^{(j)} \right] + \int_0^\infty dv' \mu_m(v')^{(k,e)} i_\omega(0,v')^{(e)}. \quad (54)$$

It contains currents due to the $(k-1)$ voltages and $(k-1)$ grids through which the beam has passed and a current due to cathode fluctuations. If there are no electrons in the tube with negative velocities, the latter current can also be expressed as a matrix produce in the form $\mathbf{\mu}_m^{(k,0)} \mathbf{i}_\omega(0)^{(e)}$, where $\mathbf{\mu}_m^{(k,0)}$ is the transmodulation matrix from the cathode plane to region (k).

XI. EVALUATION

In the Llewellyn approximation, which assumes that the second and all higher velocity moments are zero, the theory takes the form of three equations, one for the driving current i_m (which is, in fact, Eq. (10)), an equation for the density modulation $i_\omega(x)$ at the plane x (which is Eq. (4)), and an equation giving the velocity modulation $v_\omega(x)$ at that plane:

$$\begin{aligned} i_\omega &= -Y_B V_\omega + i_m, \\ i_\omega(x) &= -y_m(x) V_\omega + i_m(x), \\ -I_0 \frac{v_\omega(x)}{v(x)} &= -f_m(x) V_\omega + \left(-I_0 \frac{v_m(x)}{v(x)} \right). \end{aligned} \quad (55)$$

These are the Llewellyn electronic equations. The coefficients $y_m(x)$ and $f_m(x)$ are the density and velocity modulation admittances, respectively, and $v_m(x)$ is the velocity modulation at the plane x due to density and velocity modulations present in the beam at injection.

The second and third electronic equation can be combined as in Eq. (30) by setting

$$\mathbf{i}_\omega(x) = \begin{bmatrix} i_\omega(x) \\ -I_0 \dfrac{v_\omega(x)}{v(x)} \end{bmatrix}, \quad \mathbf{y}_m(x) = \begin{bmatrix} y_m(x) \\ f_m(x) \end{bmatrix}, \quad \mathbf{i}_m(x) = \begin{bmatrix} i_m(x) \\ -I_0 \dfrac{v_m(x)}{v(x)} \end{bmatrix}, \quad (56)$$

thus giving them a matrix representation.

The Llewellyn theory provides explicit expressions for i_m, $i_m(x)$, and $v_m(x)$:

$$\begin{aligned} i_m &= \mu_i i_\omega(0) + \mu_v \left(-I_0 \frac{v_\omega(0)}{v(0)} \right), \\ i_m(x) &= \mu_{ii}(x) i_\omega(0) + \mu_{iv}(x) \left(-I_0 \frac{v_\omega(0)}{v(0)} \right), \\ -I_0 \frac{v_m(x)}{v(x)} &= \mu_{vi}(x) i_\omega(0) + \mu_{vv}(x) \left(-I_0 \frac{v_\omega(0)}{v(0)} \right). \end{aligned} \quad (57)$$

The exciting current i_m can be written as a matrix product as in Eq. (38) by using the excitation matrix

$$\mathbf{\mu} = [\mu_i \; \mu_v], \quad (58)$$

Likewise, the second and third equations can be represented as in Eq. (36) by means of the modulation matrix

$$\mathbf{\mu}(x) = \begin{bmatrix} \mu_{ii}(x) & \mu_{iv}(x) \\ \mu_{vi}(x) & \mu_{vv}(x) \end{bmatrix}. \quad (59)$$

Thus the Llewellyn electronic equations have a

matrix representation of the form of Eqs. (60), below.

The Llewellyn calculations give the modulation admittance, excitation, and modulation matrices, as well as the beam-loading admittance, in terms of d.c. field and space-charge parameters and the d.c. transit angle.[7] The results have been used to great advantage by Llewellyn and Peterson in the discussion of vacuum-tube networks.[8] The approximation has been used by Peterson in the discussion of noise suppression in microwave tetrodes;[9] the contributions to the output noise density and velocity-modulation currents arising from cathode fluctuations can be made to cancel each other by the proper choice of the grid-screen space-charge smoothing parameter.

The Llewellyn approximation begins to lose its validity when the spread in velocities, whether thermal or due to secondaries, becomes comparable to the average velocity, and also at such frequencies that the spread leads to a spread in transit angles of an appreciable fraction of a radian. In the cathode region the theory is at best a crude approximation. However, under most operating conditions, its accuracy is quite sufficient in other regions.

Unfortunately the Llewellyn theory has no simple extension. In a region having no reflected electrons, an extension would require the calculation of the elements of the electronic equations

$$\begin{aligned} i_\omega &= -Y_B V_\omega + \mathfrak{y} \mathbf{i}_\omega(0), \\ \mathbf{i}_\omega(d) &= -\mathbf{y}_m(d) V_\omega + \mathfrak{y}(d) \mathbf{i}_\omega(0), \end{aligned} \tag{60}$$

without assuming the second and all higher velocity moments to be zero (the Llewellyn approximation). This problem has been reduced to the solution of an integral equation the kernel of which is an integral over the d.c. velocity distribution at the entrance plane of the energy derivative of a d.c. phase factor.[7] In the Llewellyn approximation the solution of this equation yields a resolving kernel which leads to the Llewellyn equations. The general solution of the integral equation would lead to Eqs. (60).

If the cathode region contains reflected electrons, the theory requires the calculation of the elements of the somewhat more general electronic equations

$$\begin{aligned} i_\omega &= -Y_B V_\omega + \int_0^\infty dv' \mu(v') i_\omega(0,v'), \\ \mathbf{i}_\omega(d) &= -\mathbf{y}_m(d) V_\omega + \int_0^\infty dv' \mathfrak{y}(d;v') i_\omega(0,v'). \end{aligned} \tag{61}$$

The reduction of this problem to an integral equation will be the subject of an investigation by the author when he returns to the study of these questions in the not too distant future. It is hoped that the formalism of this paper will facilitate study in this field, both experimental and theoretical. Many extensions suggest themselves, in particular, negative-velocity secondary currents, double injection and two-way transmission, less simple geometries, non-linear trajectories, each of which presents its own difficulties.

The Eqs. (61) are a shunt formulation in which the driving current and the current operand are expressed in terms of the r-f voltage, the exciting current, and the modulation current operand. A series formulation is also possible. Let C be the capacitance of the empty gap, $I_\omega = i_\omega - j\omega C V_\omega$ the total current, and $Y_g = j\omega C + Y_B$ the gap admittance. Then

$$\begin{aligned} V_\omega &= -Z_g I_\omega + \int_0^\infty dv' z_g(v') i_\omega(0,v'), \\ \mathbf{i}_\omega(d) &= \boldsymbol{\alpha}(d) I_\omega + \int_0^\infty dv' \boldsymbol{\alpha}(d;v') i_\omega(0,v'), \end{aligned} \tag{62}$$

is a formulation in which the r-f voltage and the current operand are expressed in terms of the total current, an exciting voltage, and a series current operand. Equations (61) and (62) are equivalent, for

$$\begin{aligned} &Z_g Y_g = 1, \quad z_g(v') = \mu(v') Z_g, \\ &\boldsymbol{\alpha}(d) = -\mathbf{y}_m(d) Z_g, \quad \boldsymbol{\alpha}(d;v') \\ &\qquad = -\mathbf{y}_m(d) z_g(v') + \mathfrak{y}(d;v'). \end{aligned} \tag{63}$$

It is perhaps well to point out that a matrix representation can be introduced for the beam even where there are electrons with negative velocities if care is taken to distinguish between positive and negative currents. Thus Eq. (60) can be used in a cathode region containing reflected electrons if only entering electrons be used in calculating the current operand $\mathbf{i}_\omega(0)$.

[7] D. R. Hamilton, J. K. Knipp, and J. B. H. Kuper, *Klystrons and Microwave Triodes* (McGraw-Hill Book Company, Inc., New York, 1948).

[8] F. B. Llewellyn and L. C. Peterson, Proc. I.R.E. **32**, 144 (1944).

[9] L. C. Peterson, Proc. I.R.E. **35**, 1264 (1947).

Barkhausen Noise and Magnetic Amplifiers. I. Theory of Magnetic Amplifiers*

J. A. KRUMHANSL** AND R. T. BEYER
Brown University, Providence, Rhode Island
(Received July 27, 1948)

An analysis is made of the operation of magnetic amplifiers or non-linear transformers. Open circuit output voltages are calculated for transformers driven by a sinusoidal primary current, and also for sinusoidal primary current plus a direct current bias. The case of a loaded secondary is also considered and it is shown that there exist conditions for instability or resonance. A subsequent paper will consider Barkhausen noise as a limiting condition on the sensitivity of these devices.

I. INTRODUCTION

THE term magnetic amplifier or non-linear transformer is customarily applied to a transformer containing ferromagnetic core material which, under the operating conditions, undergoes such magnetization that the flux density, B,[1] is not a linear function of the magnetic field intensity H. In the course of some experimental work previously reported,[2] it was noticed that the background noise which ultimately limited the sensitivity of the devices could not be entirely accounted for by the associated circuits. It seemed that this noise might be due, at least in part, to Barkhausen effect in the core materials.

In order to discuss this subject this paper (I) presents a theory of operation of these magnetic amplifiers.[3] A subsequent paper (II) will present an analysis of the Barkhausen noise in terms of a simplified physical model.

II. METHOD AND ASSUMPTIONS

The usual mode of operation of the magnetic amplifiers can be understood with the aid of Fig. 1. Two identical ferromagnetic transformers have their primary windings (N_1) connected in series. The transformers are excited by an audio oscillator which drives the cores into the region of saturation during each cycle, thus generating a potential across each secondary winding which contains the fundamental frequency plus odd harmonics. If a given secondary winding (N_2) on each transformer is connected in series opposing (relative to the primary), these signals will cancel out. However, if a direct current is entered on one of the secondaries (N_3), even harmonics will be produced across each secondary. Because of the method of connection, the contribution from each transformer will add, so that the voltage across the No. 2 windings will contain only even harmonics.

The analysis in this paper will be based on one such transformer. In the analysis it is assumed that the relation $B=F(H)$ is known, that there is no flux leakage, and that B is uniform over any cross section of the core. In most of the discussion, the effect of hysteresis is neglected. Finally, only the steady state is considered.

III. EXPRESSIONS FOR INDUCED VOLTAGES

If the constant cross-sectional area of the core is A, the magnetic flux through this area is BA. Around this core there may be a number of windings (W). Let the jth of these have N_j turns. If the current in the jth winding is I_j, the total mmf, M, is

$$M=\sum_{j=1}^{W} 0.4\pi N_j I_j, \tag{1}$$

where the summation is taken over all windings. Also, $H=M/l$, where l is the length of the core and H is the magnetic field intensity.

The voltage appearing across the terminals of the jth coil is

$$V_j=N_j\times 10^{-8}d\phi/dt, \tag{2}$$

so

$$V_j=AN_j\times 10^{-8}\left(\frac{dB}{dH}\right)_{H=M/l}\frac{d}{dt}\left(\frac{M}{l}\right). \tag{3}$$

If in each mesh there is a generator with e.m.f. E_j and linear impedance $Z_j(p)$ (where p is the operator d/dt), then the jth equation may be written

$$E_j=Z_j(p)I_j+N_j g(I_1, I_2, \cdots, I_W)\sum_{k=1}^{W} N_k p I_k, \tag{4}$$

where

$$g(I_1, I_2, \cdots, I_W)=\frac{0.4\pi A}{l}\times 10^{-8}\left(\frac{dB}{dH}\right)_{H=M/l}.$$

All voltages and currents may be expressed as harmonics of some basic frequency ω. Accordingly

* Part of the research reported in this paper was performed under contract NOrd 508. The authors wish to express their appreciation to Professor T. B. Brown for his encouragement of the work done under this contract.

** Now at Cornell University, Ithaca, New York.

[1] A table of chief symbols used in this paper is included in Appendix B.

[2] Sack, Beyer, Miller, and Trischka, Proc. I.R.E. **35**, 1375 (1947).

[3] The authors are indebted to the Bell Telephone Laboratories, whose unpublished work leading to results similar to those of Part I was made available in the course of the work mentioned in reference 2.

one can express any of these in terms of a Fourier series. Substitution in Eq. (4) and collection of coefficients for each harmonic would yield $2W$ simultaneous non-linear algebraic equations for these Fourier coefficients. Because of the tedious nature of such an operation, the discussion is limited to certain special cases, for which the equations can be "linearized."[4]

Under the assumption that $I_1 \gg I_{j \neq 1}$, the function $g(I_1, I_2, \cdots, I_W)$ can be expanded about the point P: $I_1, 0, 0, \cdots$. That is,

$$g(I_1, I_2, \cdots, I_W) = g(I_1)_P + \sum_{j=2}^{W} \left(\frac{\partial g}{\partial I_j}\right)_P I_j + \cdots. \quad (5)$$

In the above expression, $g(I_1)_P$ and each coefficient in the sum are functions of time.[5]

Then

$$g(I_1, I_2, \cdots, I_W) \sum_{k=1}^{W} N_k p I_k = g(I_1)_P N_1 p I_1 + (N_1 p I_1) \sum_{k=2}^{W} \left(\frac{\partial g}{\partial I_k}\right)_P I_k + g(I_1)_P \sum_{k=2}^{W} N_k p I_k + \text{higher order terms.} \quad (6)$$

The first step in approximating solutions to Eq. (4) is to solve

$$E_j = Z_j(p) I_j + N_1 N_j g(I_1)_P p I_1. \quad (7)$$

In this case, one can prescribe either the E_j or the I_j and solve for the other. Since Eq. (7) gives E_j explicitly, but I_j only implicitly, the I_j are prescribed.

IV. FIRST APPROXIMATIONS FOR INDUCED VOLTAGES

It shall be assumed throughout this analysis that $I_1 = I_{10} \sin \omega t$.[6] Also, the analysis is carried out for one possible approximation to the average magnetization characteristic of the form

$$B = \frac{2S}{\pi} \tan^{-1}\left(\frac{\pi \mu_0}{2S}\right) H = \frac{2S}{\pi} \tan^{-1}(\alpha H), \quad (8)$$

S is the saturation value of B and μ_0 is the initial permeability.

[4] It is often possible to carry the analysis from this point through a Fourier calculation (such as that done in the next section) without the linearizing approximation. This is done in the work cited in reference 3.

[5] Although the present discussion is concerned principally with sinusoidal excitation current, this same approximation is equally useful for other types of operating conditions. For many $B-H$ characteristics, $(\partial g/\partial I_j)_P$ is a sharply peaked function of I_1 in the neighborhood of certain values of I_1 and small elsewhere, which can simplify approximate evaluations of Fourier coefficients.

[6] While this is not the usual experimental arrangement, it serves as a useful first approximation to what is otherwise a very involved analysis.

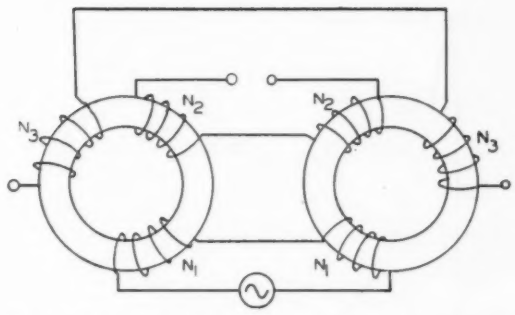

FIG. 1. Schematic diagram of windings on a pair of non-linear transformers.

The non-linear term in Eq. (7) is now expanded in a Fourier series. After some calculation, it can be shown that the coefficients $G_n(\cos)$ and $H_n(\sin)$ are given by

$$\begin{matrix} G_n \\ H_n \end{matrix} = \frac{2}{T} \int_{-T/2}^{T/2} \begin{matrix} \cos \\ \sin \end{matrix} (n\omega t) \frac{\gamma \cos \omega t \, dt}{1 + \alpha^2 D^2 \sin^2 \omega t}, \quad (9)$$

where

$$\gamma = \frac{0.4\pi A \mu_0}{l} \times 10^{-8},$$

$$D = \frac{0.4\pi}{l} N_1 I_{10}. \quad (10)$$

The solution of integrals of this type is given in Appendix A. The results here are

$$G_n = \frac{2\gamma}{\alpha D} \exp\left(-n \sinh^{-1} \frac{1}{\alpha D}\right), \quad n \text{ odd}$$

$$= 0, \quad n \text{ even or zero}$$

$$H_n = 0, \quad \text{all } n, \quad (11)$$

and the final expression for the open circuit voltage appearing across the jth winding is

$$e_j = 2 I_{10} N_1 N_j \frac{\omega \gamma}{\alpha D} \sum_{n=1,3,5} \times \exp\left(-n \sinh^{-1} \frac{1}{\alpha D}\right) \cos n\omega t. \quad (12)$$

It can easily be seen that as $D \to 0$ (the linear case), the higher harmonic voltages vanish. For $j = 1$, the coefficient of $I_{10} \cos \omega t$ becomes the linear inductance of the winding.

The relative magnitudes of the various harmonics of the open-circuit voltage appearing across the jth winding are plotted in Fig. 2. This graph indicates

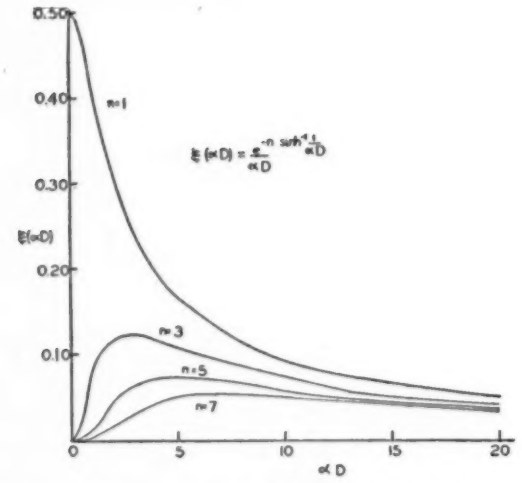

FIG. 2. Relative amplitudes of odd harmonics produced in non-linear transformers.

that there is an optimum driving condition for the production of any given harmonic.

V. FURTHER APPROXIMATIONS FOR INDUCED VOLTAGES

To consider the effects produced by currents other than I_1, Eqs. (4) and (6) are combined to give

$$E_j=Z_j(p)I_j+N_jN_1g(I_1)_Pp I_1+N_jg(I_1)_P \times\sum_{k=2}^{W}N_kpI_k+N_j(pI_1)\sum_{k=2}^{W}\left(\frac{\partial g}{\partial I_k}\right)_P I_k+\cdots. \quad (13)$$

When the same form for the magnetization curve and primary current are used as in the preceding section, it can be shown that

$$\left(\frac{\partial g}{\partial I_k}\right)_P=\frac{N_k}{N_1}\left(\frac{\partial g}{\partial I_1}\right)_P. \quad (14)$$

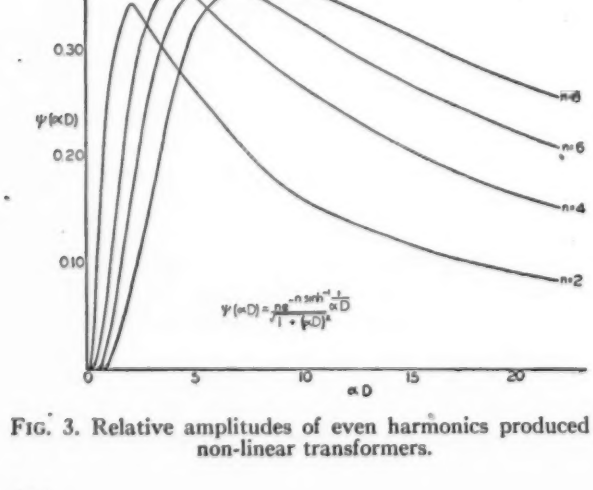

FIG. 3. Relative amplitudes of even harmonics produced in non-linear transformers.

Then Eq. (13) can be rewritten as

$$E_j=Z_j(p)I_j+N_jN_1g(I_1)_Pp I_1+N_jg(I_1)_P \times\sum_{k=2}^{W}N_kpI_k+(pI_1)\left(\frac{\partial g}{\partial I_j}\right)_P\sum_{k=2}^{W}N_kI_k. \quad (15)$$

The various terms of this expression are then developed in Fourier Series. If

$$I_j=\tfrac{1}{2}P_{j0}+\sum_{n=1}^{\infty}(P_{jn}\cos n\omega t+Q_{jn}\sin n\omega t), \quad (16)$$

and if the linear impedances at each harmonic are given by R_{jn}, X_{jn}, then

$$E_j=\tfrac{1}{2}R_{j0}P_{j0}+\sum_{n=1}^{\infty}[(R_{jn}P_{jn}+X_{jn}Q_{jn})\cos n\omega t +(R_{jn}Q_{jn}-X_{jn}P_{jn})\sin n\omega t]+N_jN_1\sum_{n=1,3,5,\cdots}^{\infty} \times\frac{2\omega\gamma}{\alpha D}I_{10}\exp\left(-n\sinh^{-1}\frac{1}{\alpha D}\right)\cos n\omega t +N_j\left[\frac{\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}}+\sum_{n=2,4,6,\cdots}^{\infty} \times\frac{2\gamma\exp\left(-n\sinh^{-1}\frac{1}{\alpha D}\right)}{(1+\alpha^2D^2)^{\frac{1}{2}}}\cos n\omega t\right] \times\left[\sum_{k=2}^{W}\sum_{n=1}^{\infty}N_kn\omega(-P_{kn}\sin n\omega t +Q_{kn}\cos n\omega t)\right]+N_j\left[-\sum_{n=2,4,6,\cdots}^{\infty} \times\frac{2n\omega\gamma\exp\left(-n\sinh^{-1}\frac{1}{\alpha D}\right)}{(1+\alpha^2D^2)^{\frac{1}{2}}}\sin n\omega t\right] \times\left[\sum_{k=2}^{W}\sum_{n=1}^{\infty}N_k(\tfrac{1}{2}P_{k0}+P_{kn}\cos n\omega t +Q_{kn}\sin n\omega t)\right]. \quad (17)$$

For one particularly simple special case, it is assumed that $R_{10}=0$, i.e., no d.c. in primary; $I_2=C=\frac{1}{2}P_{20}$ (d.c. on No. 2 winding, $P_{2n}=Q_{2n}=0$. $n\neq 0$); $W=2$ (only two windings). In this case, substitution in Eq. (17) leads to an expression for the even harmonic voltage appearing across the No.

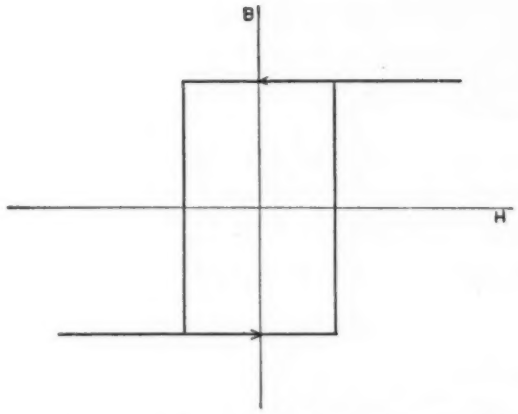

FIG. 4. Idealized $B-H$ curve used in estimating effect of hysteresis.

2 winding:

$$e_{2n}=\left[\frac{-2N_2^2\gamma n\omega\exp\left(-n\sinh^{-1}\frac{1}{\alpha D}\right)}{(1+\alpha^2D^2)^{\frac{1}{2}}}\times\sin n\omega t\right]C,\quad n=2,4,6,\cdots. \tag{18}$$

The amplitude of the e.m.f., e_{2n}, is thus directly proportional to the input direct current (for small C) as has been observed experimentally.[2] The dependence of these voltages on αD is shown in Fig. 3.

The effect of hysteresis can be studied by assuming a hysteresis loop as shown in Fig. 4. In such a case, dB/dH is represented by a delta-function. If the analysis is carried out as above, the first approximation indicates that the only effect of hysteresis is to introduce a phase shift in the output, without diminishing the amplitudes of the various harmonics. These amplitudes are affected by second-order terms, i.e., large direct currents.

In the next important case it is supposed that I_1 is prescribed but that I_2 depends on the loading conditions for No. 2 winding. As before, $I_1=I_{10}\sin\omega t$, $R_{10}=0$, and $W=2$. Under these conditions Eq. (13) reduces to

$$E_j=Z_j(p)I_j+N_jN_1g_Fp I_1+N_jN_2p(g_FI_2). \tag{19}$$

It should be noted that g_F is an even function of I_1 and pI_1 is an odd function of I_1. In order to eliminate the effects of this large second term in Eq. (19), it is customary to connect the secondaries of two identical transformers in series opposing (relative to the primaries). The total voltage across the jth set of secondaries is then

$$E_j=Z_j(p)I_j+2N_jN_2p(g_FI_2). \tag{20}$$

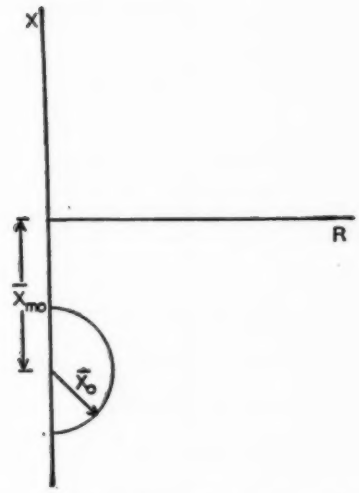

FIG. 5. Region of instability for tuned secondary on nonlinear transformers.

If the only appreciable currents in the No. 2 secondary are the d.c. and the mth harmonic, then

$$\begin{aligned}E_2=&R_{20}P_{20}+(Q_{2m}R_{2m}-P_{2m}X_{2m})\sin m\omega t\\&+(Q_{2m}X_{2m}+P_{2m}R_{2m})\cos m\omega t+2N_2^2m\omega\\&\times\left[\frac{-P_{2m}\gamma\left\{\exp\left(-2m\sinh^{-1}\frac{1}{\alpha D}\right)+1\right\}}{(1+\alpha^2D^2)^{\frac{1}{2}}}\right.\\&\left.-\frac{2P_{20}\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}}\exp\left(-m\sinh^{-1}\frac{1}{\alpha D}\right)\right]\\&\times\sin m\omega t+2N_2^2m\omega\left[\frac{-Q_{2m}\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}}\right.\\&\left.\times\left\{\exp\left(-2m\sinh^{-1}\frac{1}{\alpha D}\right)-1\right\}\right]\cos m\omega t.\end{aligned} \tag{21}$$

If the coefficients of $\sin m\omega t$, $\cos m\omega t$, and d.c. term are equated, and if the notation

$$\bar{X}_m=\frac{2N_2^2m\omega\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}}\exp\left(-2m\sinh^{-1}\frac{1}{\alpha D}\right),$$

$$\bar{X}_{m0}=\frac{2N_2^2m\omega\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}},$$

$$e_{20m}=\frac{4N_2^2m\omega\gamma}{(1+\alpha^2D^2)^{\frac{1}{2}}}\left(\frac{E_2}{R_{20}}\right)\exp\left(-m\sinh^{-1}\frac{1}{\alpha D}\right) \tag{22}$$

is used, further calculation leads to

$$\begin{aligned}e_{20m}&=R_{2m}Q_{2m}-(\bar{X}_m+X_{2m}+\bar{X}_{m0})P_{2m},\\0&=(X_{2m}-\bar{X}_m+\bar{X}_{m0})Q_{2m}+R_2P_{2m}.\end{aligned} \tag{23}$$

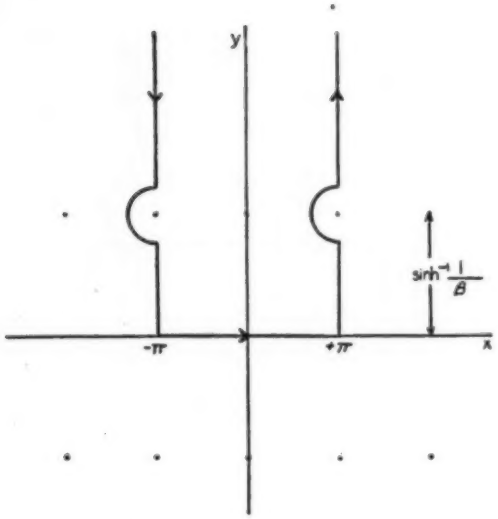

FIG. 6. Integration contour.

The current amplitudes are then

$$Q_{2m}=\frac{e_{20m}R_{2m}}{\Delta}$$

$$P_{2m}=-\frac{(X_{2m}-\bar{X}_m+\bar{X}_{m0})e_{20m}}{\Delta}, \qquad (24)$$

where

$$\Delta=R_{2m}^2-\bar{X}_m^2+(X_{2m}+\bar{X}_{m0})^2.$$

Since Δ vanishes for certain values of the parameters, it follows that there exists a region of instability, the locus of which is the semicircle drawn in Fig. 5.

One should note that apparently the behavior in the vicinity of the mth harmonic is analogous to a voltage generator which supplies the "open circuit" voltage e_{20m} to a four-terminal network whose transfer characteristics are described by Eq. (21).

APPENDIX A

The various Fourier expansions involve integrals of the form

$$I(K,\beta)=\int_{-\pi}^{\pi}\frac{e^{iKz}dz}{1+\beta^2\sin^2 z}. \qquad (25)$$

In the corresponding contour integral, the poles are given by

$$z=\pm i\sinh^{-1}\frac{1}{\beta}+n\pi, \quad n=0,\pm1,\pm2,\cdots. \qquad (26)$$

The contour for positive K is shown in Fig. 6. Since the integrand $\to 0$ as $z\to x+i\infty$, the upper side of the contour does not contribute to the integral. Also, the opposite sides differ in value of z by 2π so that their contributions cancel. Hence the value of the contour integral is the same as that of Eq. (25).

The residue at $z=i\sinh^{-1}1/\beta$ is found to be

$$R_0=\frac{\exp\left(-K\sinh^{-1}\frac{1}{\beta}\right)}{2(1+\beta^2)^{\frac{1}{2}}}, \qquad (27)$$

while that at $z=i\sinh^{-1}1/\beta+\pi$ is

$$R_0=(-1)^K\frac{\exp\left(-K\sinh^{-1}\frac{1}{\beta}\right)}{2(1+\beta^2)^{\frac{1}{2}}}. \qquad (28)$$

The value of the integral is then

$$I(K,\beta)=\frac{2\pi}{(1+\beta^2)^{\frac{1}{2}}}\exp\left(-K\sinh^{-1}\frac{1}{\beta}\right), \quad K \text{ even}$$

$$=0, \quad K \text{ odd}. \qquad (29)$$

For negative K, the integration is performed in the lower half plane and Eq. (29) is again obtained, except that $|K|$ replaces K.

APPENDIX B

Table of Symbols

B	Flux density.
H	Magnetic field intensity.
M	Magnetomotive force.
S	Saturation value of flux density.
ϕ	Flux.
A	Cross-sectional area of core.
l	Length of core.
W	Number of windings on core.
N_j	Number of turns on jth winding.
I_j	Current in jth winding.
E_j	E.m.f. in jth mesh.
$Z_j(p)$	Linear impedance in jth mesh.
p	d/dt.
R_{jn}	Resistance in jth mesh at nth harmonic.
X_{jn}	Reactance of load impedance of jth winding at nth harmonic.
ω	Angular frequency of driving oscillator.
μ_0	Initial permeability of core.
α	$\pi\mu_0/2S$.
C	Direct current input in No. 2 winding.
e_{2n}	Even harmonic voltages produced in No. 2 winding.
I_{10}	Amplitude of current in No. 1 winding.
$g(I_1, I_2, \cdots, I_W)$	$\gamma(1+\alpha^2D^2\sin^2\omega t)^{-1}$.
$g(I_1)_P$	$g_P(I_1, 0, \cdots, 0)$.
γ	$(0.4\pi A/l)\mu_0\times10^{-8}$.
D	$(0.4\pi/l)N_1I_{10}$.
P_{jn}, Q_{jn}	Fourier coefficients of expansion for I_j.
G_n, H_n	Fourier coefficients of expansion for $g(I_{10}\sin\omega t)\cos\omega t$.

On the Infinitely Long Cylindrical Antenna*

CHARLES H. PAPAS
Cruft Laboratory, Harvard University Cambridge, Massachusetts
(Received October 13, 1948)

Using the method of steepest descents the far field and the asymptotic form of the current distribution is obtained for an infinitely long, perfectly conducting cylindrical antenna excited by a localized electromotive force. The low frequency value of the radiation conductance is determined by integrating the radiated energy flux over a large sphere.

I. INTRODUCTION

THE problem of determining the longitudinal distribution of current and the input impedance of an infinitely long, perfectly conducting circular cylinder antenna has interested several investigators. Among them are Stratton and Chu,[1] Schelkunoff,[2] and Hallén.[3] And recently a new method of analyzing the problem was brought to the author's attention through an unpublished note by H. Levine in which he elegantly formulates the input impedance by means of the free-space dyadic Green's function and a variational principle involving the current. Using a simple plane wave-trial function, he obtains an expression for the input impedance which is valid for a broader frequency range than the expressions of the aforementioned investigators.

Unless one uses the variational approach, the core of the problem lies in the evaluation of a certain contour integral. Once this integral has been evaluated the current at any point along the antenna and hence its input impedance is readily determined. In this paper, by means of the method of steepest descents we evaluate this integral for distances far from the localized electromotive force exciting the antenna. This gives the explicit form of the far-zone field from which the asymptotic value of the antenna current is easily calculated. Moreover, the input conductance of the antenna is obtained by integrating the radiated energy flux over a large sphere.

To formulate the problem we assume that an infinitely long, perfectly conducting circular cylinder of radius a is driven at some cross section by a localized electromotive force, V, circumscribing the cylinder in a narrow peripheral band. Since only longitudinal currents can be generated by such a source, the axial and radial components of the magnetic vector H and the circumferential component of the electric vector E are identically zero. That is,

$$E=(E_\rho, 0, E_z) \tag{1}$$

$$H=(0, H_\phi, 0) \tag{2}$$

where E_ρ, H_ϕ, E_z are the non-vanishing components of the field in cylindrical coordinates (Fig. 1). The exact expression for $E_z(\rho, \phi, z)$ is given by

$$E_z(\rho, \phi, z)=\frac{V}{2\pi}\int_{-\infty}^{\infty}\frac{H_0^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}\rho]}{H_0^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}a]}e^{i\zeta z}d\zeta \tag{3}$$

where $H_0^{(1)}(x)$ is the Hankel function of the first kind and zero order, $k=2\pi/\lambda$, λ being the free-space wave-length, and ζ is a complex variable. The path of integration is along the real axis of the complex ζ-plane with an upward indentation at $\zeta=-k$ and a downward one at $\zeta=+k$ (Fig. 2). This equation can be obtained from previous investigations,[1,2,3] or from a report by Papas[4] in which he derives integral expressions for the field components and the antenna current by means of a Green's function used by Levine and Schwinger.[5] Our first task is to find the explicit form of Eq. (3) for large values of r.

II. FAR FIELD

To find the far field, i.e., find E_z when r is large, the method of steepest descents[6] is used. That is, the integral

$$E_z(\rho, \phi, z)=\frac{V}{2\pi}\int_{-\infty}^{\infty}\frac{H_0^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}\rho]}{H_0^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}a]}e^{i\zeta z}d\zeta \tag{3}$$

is evaluated for large r. Let $u=(k^2-\zeta^2)^{\frac{1}{2}}$; then for ρ large and $u\neq 0$,

$$H_0^{(1)}(u\rho)e^{i\zeta z}\sim\left\{\frac{2}{\pi u\rho}\right\}^{\frac{1}{2}}\exp\left[i\left(u\rho-\frac{\pi}{4}+\zeta z\right)\right]. \tag{4}$$

* The research reported in this paper was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5ori-76, T.O.I.

[1] J. A. Stratton and L. J. Chu, J. App. Phys. **12**, 230–235 (1945).

[2] S. A. Schelkunoff, Proc. I.R.E. **29**, 493–521 (1941); Proc. I.R.E. **33**, 872–878 (1945).

[3] E. Hallén, J. App. Phys. **19**, 1140–1147 (1948).

[4] C. H. Papas, "On the Infinitely Long Cylindrical Antenna," Cruft Laboratory Technical Report No. 58, Harvard University, September 10, 1948.

[5] H. Levine and J. Schwinger, Phys. Rev. **73**, 383 (1948).

[6] P. Debye, Math. Ann. **LXVII**, 535–558 (1909); "Semikonvergente Entwickelungen fur die Zylinderfunktionen und ihre Ausdehung ins Komplexe," Munchener Sitzungsberichte, **XL** (1910).

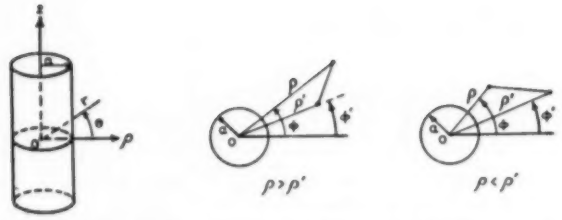

FIG. 1. Coordinate systems used in the formulation of the problem. Cylindrical coordinates: ρ, ϕ, z. Spherical coordinates: r, θ, ϕ.

The value of ζ which makes the exponential $e^{i(u\rho+\zeta z)}$ an extremum, is a saddlepoint and is denoted by ζ_0. Hence ζ_0 is the solution of

$$\frac{d}{d\zeta}[(k^2-\zeta^2)^{\frac{1}{2}}\rho+\zeta z]=0.$$

Carrying out the differentiation, it is easily found that

$$\zeta_0=\frac{kz}{(\rho^2+z^2)^{\frac{1}{2}}}=k\sin\theta \tag{5}$$

where θ is measured from the plane $z=0$. At $\zeta=\zeta_0$

$$e^{i(u\rho+\zeta z)}=e^{ik(\rho\cos\theta+z\sin\theta)}$$

and in the neighborhood of ζ_0 it is approximately equal to

$$\exp\left[ik(\rho\cos\theta+z\sin\theta)-\frac{i\rho}{2k\cos^3\theta}(\zeta-\zeta_0)^2\right]. \tag{6}$$

The path of integration through the saddlepoint is determined by the constancy of the imaginary part of the exponent of Eq. (6); that is, the path is so chosen that the imaginary part of the exponent remains constant as the path of integration traverses the saddlepoint. Since the exponent is a monogenic function and its imaginary part remains constant, its real part rises from negligibly small values to a peak value at the saddlepoint and then down again. The requirement for the constancy of the imaginary part of the exponent is expressed by

$$\text{I.P.}[ik(\rho\cos\theta+z\sin\theta)]$$

$$=\text{I.P.}\left[ik(\rho\cos\theta+z\sin\theta)-\frac{i\rho(\zeta-\zeta_0)^2}{2k\cos^3\theta}\right]. \tag{7}$$

Let $\zeta-\zeta_0=se^{i\alpha}$. In this, α is the angle which the path makes with the real axis of the ζ-plane. Substituting this into Eq. (7), it is found that if (7) is to be satisfied, it is necessary that the R.P. $(s^2e^{2i\alpha})=0$; i.e., $\cos 2\alpha=0$ or $\alpha=\pm\pi/4$. The value $\alpha=\pi/4$ is excluded because the path through ζ_0 must extend from the second to the fourth quadrant of the complex ζ-plane; this leaves $\alpha=-\pi/4$ as the angle which the path of steepest descent makes with the real axis of the complex ζ-plane.

We now transform[7] from the ζ-plane to the complex τ-plane by the following transformation

$$\zeta=k\sin\tau$$
$$\tau=\theta+i\psi. \tag{8}$$

The path of integration in the ζ-plane (Fig. 2) transforms into path C in the τ-plane of Fig. 3. The saddlepoints lie on the real axis of the τ-plane, i.e., $\tau=\theta$ is a saddlepoint. We de-form C into C_0, which starts at $-(\pi/2)+i\infty$, goes through $\theta=0$, and then ends at $\pi/2+i\infty$. A path such as C_θ passing through a saddlepoint θ would have to start at $\theta-(\pi/2)+i\infty$ and terminate at $\theta+(\pi/2)+i\infty$. In the τ-plane, Eq. (4) transforms into

$$\left(\frac{2}{\pi\rho k\cos\tau}\right)^{\frac{1}{2}}e^{-i(\pi/4)}e^{ik(\rho\cos\tau+z\sin\tau)}$$

$$=\left(\frac{2}{\pi\rho k\cos\tau}\right)^{\frac{1}{2}}e^{-i(\pi/4)}e^{ikr\cos(\tau-\theta)}. \tag{9}$$

Let $F(k\cos\tau)$ represent the factor

$$\{H_0^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}a]\}^{-1}$$

in the τ-plane. Substituting this and Eq. (9) into (3), we get

$$E_z(\rho,z)=\frac{V}{\pi}e^{-i(\pi/4)}\int_{\theta-(\pi/2)+i\infty}^{\theta+(\pi/2)-i\infty}\left\{\frac{2}{\pi\rho k\cos\tau}\right\}^{\frac{1}{2}}$$
$$\times e^{ikr\cos(\tau-\theta)}F(k\cos\tau)k\cos\tau d\tau. \tag{10}$$

This integration is carried out along the path C_θ. It is only necessary to carry out the integration on a short segment of the path in the neighborhood of $\tau=\theta$. We set $\tau-\theta=\eta e^{-i(\pi/4)}$ where η is the distance along the path measured from the saddlepoint. When $|\tau-\theta|$ is small,

$$\cos(\tau-\theta)\doteq 1-\frac{(\tau-\theta)^2}{2}=1-\frac{\eta^2}{2}e^{-i(\pi/2)}.$$

Moreover, $d\tau=d\eta e^{-i(\pi/4)}$. With these, Eq. (10) takes the form

$$E_z(\rho,z)=-\frac{Vi}{\pi}e^{ikr}F(k\cos\theta)$$
$$\times\left\{\frac{2k\cos\theta}{\pi\rho}\right\}^{\frac{1}{2}}\int_{-\epsilon}^{\epsilon}e^{-(kr\eta^2/2)}d\eta. \tag{11}$$

Since r is large we can replace the limits, $-\epsilon$ to ϵ,

[7] F. Noether, "Elektromagnetische Wellen an einem Draht, bei konzentrierter Energiequelle;" Physik Zeits. USSR, Band 8, Heft 1, 1–24 (1935).

by infinite ones. That is,

$$\int_{-\epsilon}^{\epsilon} e^{-(kr\eta^2/2)}d\eta \doteq \int_{-\infty}^{\infty} e^{-(kr\eta^2/2)}d\eta = \left(\frac{2\pi}{kr}\right)^{\frac{1}{2}}. \quad (12)$$

Recalling that $r\cos\theta = \rho$, Eq. (11) takes the form

$$E_z(r,\theta) = -\frac{Vi}{\pi}\frac{e^{ikr}}{r}\frac{1}{H_0^{(1)}(ka\cos\theta)}. \quad (13)$$

This is the far field of an infinitely long antenna driven by a localized e.m.f. V. The θ-component of the E-field and the ϕ-component of the H-field are given respectively by

$$E_\theta(r,\theta) = \frac{E_z(r,\theta)}{\cos\theta} = -\frac{Vi}{\pi}\frac{e^{ikr}}{r\cos\theta}\frac{1}{H_0^{(1)}(ka\cos\theta)} \quad (14)$$

and

$$H_\phi(r,\theta) = \frac{E_\theta(r,\theta)}{120\pi} = -\frac{Vi}{120\pi^2}\frac{e^{ikr}}{r\cos\theta} \times \frac{1}{H_0^{(1)}(ka\cos\theta)}. \quad (15)$$

III. THE RADIATION CONDUCTANCE

To find the radiation conductance of the antenna, the radial component of Poynting's vector S must be integrated over the surface of a large sphere. This integration yields the radiated power P, and division by $V^2/2$ gives the radiation conductance.

Poynting's vector is given by

$$S = \frac{1}{2}E_\theta H_\phi^* = \frac{V^2}{240\pi^3 r^2\cos^2\theta} \times \frac{1}{[H_0^{(1)}(ka\cos\theta)][H_0^{(1)}(ka\cos\theta)]^*} \quad (16)$$

where the asterisk indicates conjugate complex values. Thus

$$P = \int_{-\pi/2}^{\pi/2} 2\pi r^2 S\cos\theta d\theta = \frac{V^2}{120\pi^2}\int_{-\pi/2}^{\pi/2}\frac{1}{\cos\theta} \times \frac{d\theta}{[H_0^{(1)}(ka\cos\theta)][H_0^{(1)}(ka\cos\theta)]^*} \quad (17)$$

when ka is small,

$$H_0^{(1)}(ka\cos\theta) \to 1 - \frac{2i}{\pi}\log\frac{2}{\gamma ka\cos\theta},$$

and the integral in Eq. (17) is approximately

FIG. 2. Paths of integration in the ζ-plane.

equal to

$$2\int_0^{\pi/2}\frac{d\theta}{\cos\theta\left|1-\frac{2i}{\pi}\log\frac{2}{\gamma ka\cos\theta}\right|^2} = 2\int_0^{\pi/2} \times \frac{d\theta}{\cos\theta\left|1-\frac{2i}{\pi}\left(\log\frac{2}{\gamma}-\log ka-\log\cos\theta\right)\right|^2} \quad (18)$$

where $\log 2/\gamma = 0.1159$. This integral appears to be rather difficult to handle rigorously, and we therefore resort to an approximation.

Let us examine the integrand

$$\frac{1}{\cos\theta\left|1-\frac{2i}{\pi}\left(\log\frac{2}{\gamma}-\log ka-\log\cos\theta\right)\right|^2}. \quad (19)$$

Since ka is small, the term in square brackets can be replaced by $(4/\pi^2)(\log ka+\log\cos\theta)^2$. This quantity varies from $(4/\pi^2)(\log ka)^2$ to $(16/\pi^2)(\log ka)^2$ as θ varies from $\theta=0$ to $\theta=\theta_1=\arccos ka$. For this range we replace $(4/\pi^2)(\log ka+\log\cos\theta)^2$ by the geometric mean of its end values, i.e. $(8/\pi^2)(\log ka)^2$. Hence for the range $\theta\leq\theta\leq\theta_1$, the contribution to Eq. (18) is

$$2\int_0^{\theta_1}\frac{\pi^2}{8}\frac{1}{[\log ka]^2}\frac{d\theta}{\cos\theta} = \frac{\pi^2}{4}\frac{1}{[\log ka]^2} \times \log\tan\left(\frac{\pi}{4}+\frac{\theta_1}{2}\right) \doteq \frac{\pi^2\log\operatorname{ctn}\left(\frac{ka}{2}\right)}{4[\log ka]^2}, \quad (20)$$

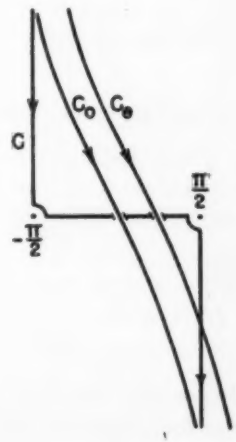

FIG. 3. Paths of integration in the τ-plane.

since $\theta_1 = \arccos ka \doteq \pi/2 - ka$. Moreover $\log \operatorname{ctn} ka/2 \doteq \log 2/ka = \log 2 + \log 1/ka$. Neglecting the term $\log 2$, (20) is approximately equal to

$$2\int_0^{\theta_1} \frac{\pi^2}{8} \frac{1}{[\log ka]^2} \frac{d\theta}{\cos\theta} \doteq \frac{\pi^2}{4} \frac{1}{\log \frac{1}{ka}}. \tag{21}$$

To complete the range of integration we again approximate the integrand Eq. (19)

$$2\int_{\theta_1}^{\pi/2} \frac{d\theta}{\cos\theta \left[\frac{2}{\pi} \log \frac{2}{\gamma ka \cos\theta}\right]^2} \doteq \frac{\pi^2}{2} \int_{\theta_1}^{\pi/2} \times \frac{d\theta}{\left(\frac{\pi}{2} - \theta\right)\left(\log\left\{ka\left(\frac{\pi}{2} - \theta\right)\right\}\right)^2}. \tag{22}$$

This integral can be immediately evaluated and has the value $(\pi^2/4)[1/\log(1/ka)]$. By adding this result to Eq. (21) the approximate value of (17) is obtained. Hence,

$$P = \frac{V^2}{120\pi^2}\left[\frac{\pi^2}{4} \frac{1}{\log \frac{1}{ka}} + \frac{\pi^2}{4} \frac{1}{\log \frac{1}{ka}}\right].$$

The radiation conductance G is given by

$$G = \frac{2P}{V^2} = \frac{1}{120 \log\left(\frac{1}{ka}\right)}. \tag{23}$$

On physical grounds G must be positive. Consequently it is necessary that $1/ka > 1$ or $\lambda > 2\pi a$. ($k = 2\pi/\lambda$ where λ is the wave-length of the exciting source). This means that Eq. (23) is a low frequency, i.e., long wave-length, approximation to the radiation conductance. When $ka = 1$, G becomes infinite.

IV. ASYMPTOTIC FORM OF THE CURRENT

For great distances from the localized e.m.f. the current distribution can be computed from the far-field expression (15). For large values of r we found that

$$H_\phi(r, \theta) = -\frac{Vi}{120\pi^2} \frac{e^{ikr}}{r} \frac{1}{\cos\theta\, H_0^{(1)}(ka\cos\theta)}. \tag{24}$$

To obtain the value of H_ϕ on the surface of the cylinder for large z, it is necessary to restrict r and θ so that $r\cos\theta = a$; moreover, $r\sin\theta = z$. But since $\theta \doteq \pi/2$, we have $r \doteq z$. The argument of the Hankel function is given by the approximation: $ka\cos\theta \doteq ka^2/z$. Substituting these into Eq. (24) we obtain H_ϕ for $\rho = a$ and z large:

$$(H_\phi)_{\substack{\rho = a \\ z \to \infty}} = -\frac{Vi}{120\pi^2} \frac{e^{ikz}}{a} \frac{1}{H_0^{(1)}\left(\frac{ka^2}{z}\right)}. \tag{25}$$

The current for large z is then given by

$$I(z) = 2\pi a H_\phi = -\frac{Vi}{60\pi} \frac{e^{ikz}}{H_0^{(1)}\left(\frac{ka^2}{z}\right)}. \tag{26}$$

Since

$$\frac{ka^2}{z} \ll 1, \quad H_0^{(1)}\left(\frac{ka^2}{z}\right) \doteq -\frac{2i}{\pi} \log \frac{2z}{\gamma ka^2}.$$

Using this approximation the asymptotic form of the current takes the form

$$I(z) = \frac{V}{120} \frac{e^{ikz}}{z \log \frac{z}{ka^2}}. \tag{27}$$

V. CONCLUSIONS

The far field (Eqs. (14), (15)) and the asymptotic form of the current at large distances from the source of excitation (Eq. (27)) were determined by the method of steepest descent. By integrating the radiated energy flux an approximate value of the radiation conductance was calculated (Eq. (23)). This integration, however, yields no information about the radiation susceptance since the sphere over which the integration is performed lies in the far zone where the time average energy flux is purely real. There is close agreement for long wave-lengths, i.e., for low frequencies, among the values of conductance obtained by Schelkunoff,[2] Hallén,[3] and Eq. (23). For shorter wave-lengths, Eq. (23) becomes invalid since the radiation conductance becomes infinite when $\lambda = 2\pi a$ and negative when $\lambda < 2\pi a$.

ACKNOWLEDGMENT

The author wishes to thank Dr. Harold Levine of Harvard University for his very generous assistance.

Note added in proof: Professor S. Silver has also derived (3). See Report No. 149, Antenna Lab., University of California, Jan. 3, 1949, "The Field of a Slot of Arbitrary Shape In an Infinite Cylinder."

Phase Contrast in Electron Microscope Images*

E. G. Ramberg
RCA Laboratories, Princeton, New Jersey
(Received November 1, 1948)

Image contrast in electron microscope images may arise from absorption, scattering, or phase changes impressed on the illuminating electron wave. In the present paper the intensity distribution in the in-focus and out-of-focus image of an edge of a transparent thin film introducing a prescribed phase change in the incident beam is calculated. It is found that the resulting "phase contrast" increases both with the film thickness (i.e., magnitude of phase change) and with the degree of defocusing and, for thin specimens, exceeds other sources of contrast in magnitude. The model is too schematic to permit a quantitative comparison with measured intensity distributions.

A THIN specimen may modify an incident electron wave by absorption, scattering, and changing the phase of the electron wave. All these may give rise to contrasts in electron microscope images. With thin specimens absorption is negligible. Furthermore, the scattering of electrons beyond the effective aperture of the objective may yield only slight contrast. At the same time, phase changes in the electron waves traversing the specimen may give rise to a pronounced delineation of specimen edges. This phase contrast, at the edge of a specimen of uniform thickness, will be investigated in the present paper. For reference, it may be recalled that a 100-angstrom-unit film of a material with an inner potential of 15 volts produces a phase delay of approximately 90° in a 50,000-volt electron beam.[1]

I. INTENSITY DISTRIBUTION IN THE IN-FOCUS AND OUT-OF-FOCUS IMAGE OF AN OPAQUE EDGE

The first step in determining the phase contrast in the image of the edge of a transparent film, whose sole effect on the incident electron wave is to introduce a prescribed phase change, is the calculation of the wave amplitude distribution in the in-focus and out-of-focus image of a perfectly opaque edge. It will be assumed that a plane electron wave is incident on an opaque half-plane a distance d_0 from the front focal point of the objective, which will be regarded as aberration-free within the limiting aperture of radius $\rho=\theta f$.[2] The geometrical arrangement is, thus, that pictured in Fig. 1. The notation employed is the following:

f: focal length of objective.
$\rho=\theta f$: radius of lens aperture.
d_0: distance of front focal plane from edge.
$d=f^2/d_0$: distance of image of edge from rear focal plane (d, d_0 positive corresponds to too strong objective current).
ξ, η, ζ: coordinates in aperture plane.
x, y, z: coordinates in reference plane.
$\alpha=x/z$

$$C(w)+iS(w)=\int_0^w \exp[\pi i x^2/2]dx;$$

$$Si(w)=\int_0^w \sin x dx/x.$$

Specifically, the amplitude distribution will be determined, for small values of α, along the x axis in a reference plane an infinite distance from the objective ($z=\infty$).

The procedure is to determine the wave amplitude $u_p{}^0$ in the (image-side) aperture plane of the lens, assuming that the latter is perfect and unrestricted and to find the excitation at the reference point (x, y, z) corresponding to the wave amplitude distribution in the actual aperture $(\xi^2+\eta^2<\rho^2)$ by means of the formula[3]

$$u_P=(1/2\pi)\int_{-\rho}^{\rho}\int_{-(\rho^2-\xi^2)^{\frac{1}{2}}}^{(\rho^2-\xi^2)^{\frac{1}{2}}}(z/r)u_P{}^0(\xi,\eta)\times\frac{\partial}{\partial\zeta}(e^{ikr}/r)d\xi d\eta, \quad (1)$$

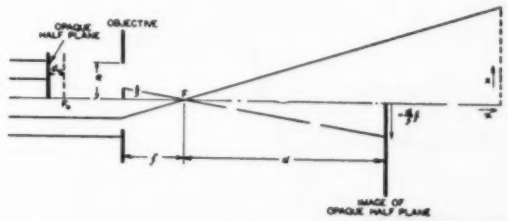

FIG. 1. Geometrical arrangement for the determination of the intensity distribution in the out-of-focus image of an opaque half-plane.

* Presented at the E. F. Burton Memorial Meeting of the Electron Microscope Society of America in Toronto on September 11, 1948.

[1] See, e.g., J. Hillier and E. G. Ramberg, J. App. Phys. 18, 48–71 (1947).

[2] This implies that the physical aperture of the objective is made appreciably smaller than the aperture which is optimum from the point of view of resolution.

[3] R. K. Luneberg, *Mathematical Theory of Optics* (Brown University, Providence, Rhode Island, 1944), p. 354. The author is indebted to Dr. M. Herzberger for pointing out Luneberg's formulation of the diffraction problem.

with

$$r=[(z-\zeta)^2+(x-\xi)^2+(y-\eta)^2]^{\frac{1}{2}}.$$

The excitation in the aperture plane is simply the complex conjugate of that produced by a spherical wave converging at the focal point F, incident from the right on a real opaque half-plane placed at the geometric image of the opaque half-plane in object space. This is given by[4]

$$u_P{}^0=(i/2)(f^2+\xi^2+\eta^2)^{-\frac{1}{2}}\exp[-ik(f^2+\xi^2+\eta^2)^{\frac{1}{2}}]$$
$$\times\{S(w)-C(w)+i[1-C(w)-S(w)]\}, \quad (2)$$

with

$$w=\xi\left\{\frac{k}{\pi}\frac{d}{f+d}\frac{f^2}{(f^2+\xi^2+\eta^2)^{\frac{1}{2}}}\right\}^{\frac{1}{2}}$$
$$\cong\xi\left(\frac{k}{\pi}\frac{d}{f(d+f)}\right)^{\frac{1}{2}}\left(1-\frac{3}{4}\frac{\xi^2+\eta^2}{f^2}\right).$$

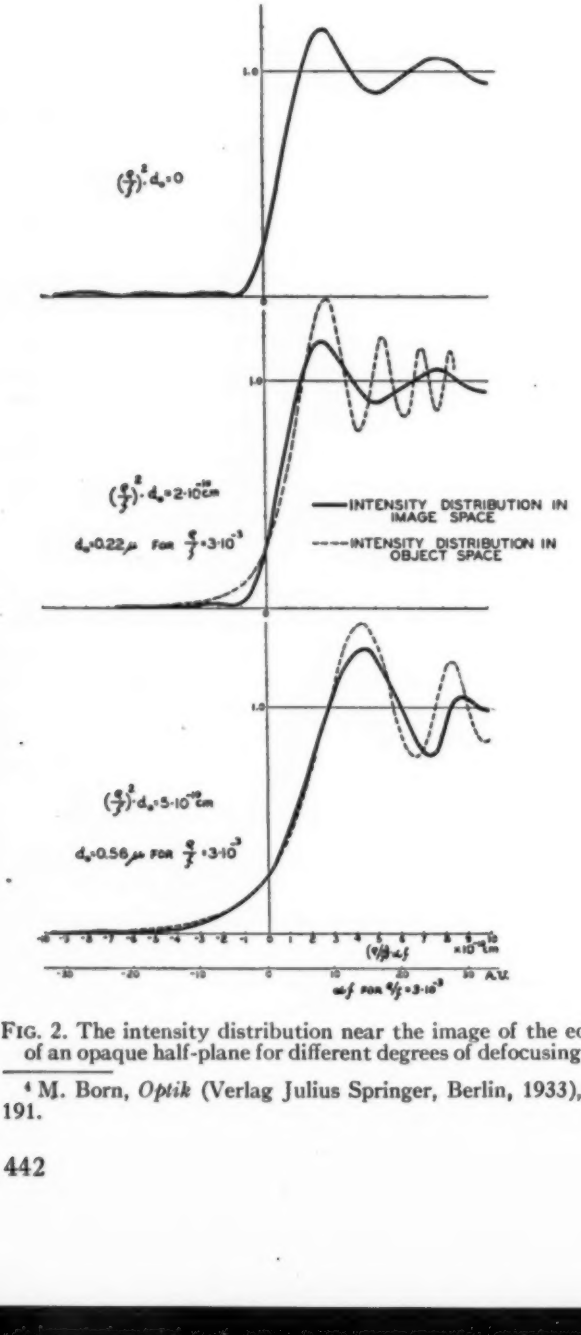

FIG. 2. The intensity distribution near the image of the edge of an opaque half-plane for different degrees of defocusing.

[4] M. Born, *Optik* (Verlag Julius Springer, Berlin, 1933), p. 191.

Equation (1) may be brought into reasonably simple form if following approximations are made:

(i) $1/r$ and $1/f$ are neglected in comparison with $k=2\pi/\lambda$ in the coefficients.
(ii) $(\xi^2+\eta^2)/(2f)$ is neglected in comparison with f in the coefficients.
(iii) terms of order higher than $(\xi^2+\eta^2)/(2f)$ are neglected in comparison with f in the exponents and arguments of the Fresnel integrals.
(iv) α^2 is neglected in comparison with 1 in the coefficients.

A check of numerical values indicates that all these approximations are justified in the present application. Then

$$u_P(x,0,z)=-\frac{i}{4\pi fz}\exp[ik(-f+z[1+\alpha^2]^{\frac{1}{2}})]$$
$$\times 2ik\int_{-\rho}^{\rho}d\xi\int_0^{(\rho^2-\xi^2)^{\frac{1}{2}}}d\eta\exp\left[-ik\left[\alpha\xi+\frac{\xi^2+\eta^2}{2f}\right]\right]$$
$$\times[S(w)-C(w)+i\{1-C(w)-S(w)\}], \quad (3)$$

with

$$w=\frac{\xi}{f}\left(\frac{kf}{\pi}\frac{d}{f+d}\right)^{\frac{1}{2}}.$$

Integration with respect to η simply replaces the factor $\exp[-k\eta^2/(2f)]d\eta$ by

$$\left(\frac{\pi f}{k}\right)^{\frac{1}{2}}\left[C\left(\left[\frac{k}{\pi f}(\rho^2-\xi^2)\right]^{\frac{1}{2}}\right)-iS\left(\left[\frac{k}{\pi f}(\rho^2-\xi^2)\right]^{\frac{1}{2}}\right)\right],$$

which for large argument (ξ not too close to ρ) is equivalent to $(\pi f/(2k))^{\frac{1}{2}}e^{-i\pi/4}$. For convenience in calculation it is thus possible to break up the integral for u_P in the following manner, making use of the fact that $C(-w)=-C(w)$, $S(-w)=-S(w)$:

$$u_P\cong-[i/(4z)](\pi kf)^{-\frac{1}{2}}\exp[ik(-f+z(1+\alpha^2)^{\frac{1}{2}})]$$
$$\times\{I_1+I_2+I_3\}. \quad (4)$$

Here

$$I_1=2k(1-i)\int_0^{\rho}d\xi\exp[-ik[-\alpha\xi+\xi^2/(2f)]]$$
$$\cong-i(4\pi kf)^{\frac{1}{2}}\{1+\alpha(fk/\pi)^{\frac{1}{2}}$$
$$\times[1-\cos(k\rho^2/2f)-\sin(k\rho^2/2f)]\} \quad (5)$$

and

$$I_2=-4ik\int_0^{\rho}\sin(k\alpha\xi)\exp[-ik\xi^2/(2f)]$$
$$\times\{[\tfrac{1}{2}-S(w)]-i[\tfrac{1}{2}-C(w)]\}d\xi$$
$$\cong-4\pi kf\alpha\int_0^{w_2}w\exp[-i\pi w^2/2]$$
$$\times\{[\tfrac{1}{2}-C(w)]+i[\tfrac{1}{2}-S(w)]\}dw$$
$$+4i\alpha kf\int_{w_2}^{(\rho/f)(kf/\pi)^{\frac{1}{2}}}\frac{\sin[(\pi kf)^{\frac{1}{2}}\alpha w]}{(\pi kf)^{\frac{1}{2}}\alpha w}$$
$$\times\left\{\cos\frac{\pi fw^2}{2d}-i\sin\frac{\pi fw^2}{2d}\right\}dw \qquad (6)$$

with, e.g., $w_2=2$. Finally,

$$I_3=-4k\int_0^{\rho}d\xi\exp[ik[-\alpha\xi+\xi^2/(2f)]]$$
$$\times\{[\tfrac{1}{2}-C(w')]-i[\tfrac{1}{2}-S(w')]\} \qquad (7)$$

with

$$w'=[k(\rho^2-\xi^2)/(\pi f)]^{\frac{1}{2}}.$$

This has the approximate value

$$I_3\cong i(4\pi kf)^{\frac{1}{2}}\exp[-ik\rho^2/(2f)]$$
$$\times\{J_0(k\alpha\rho)+iH_0(k\alpha\rho)\}. \qquad (8)$$

Here J_0 is Bessel's function and H_0 Struve's function of order zero.[5] The last integral does not depend on the location of the edge and is clearly simply the Fresnel diffraction pattern of the aperture, vanishing except near the optic axis of the system.[6] It is thus not characteristic of the pattern at the edge of an object and is omitted from the remaining considerations.

Omitting a common factor $-(1/z)\exp[ik(-f+z(1+\alpha^2)^{\frac{1}{2}})]$ we thus find, for small deviations from focus and for the region close to the geometric image of the edge, the approximate relation

$$u_P\cong\tfrac{1}{2}+(1/\pi)\int_0^{\rho k\alpha}(\sin x/x)$$
$$\times\{\cos(x^2/2\alpha^2dk)-i\sin(x^2/2\alpha^2dk)\}dx. \qquad (9)$$

For exact focus ($d\to\infty$) this becomes $\frac{1}{2}+(1/\pi)\times Si(\rho k\alpha)$. On the other hand, for large deviations from focus (i.e., $d=f^2/d_0$ small) the distribution approaches the Fresnel diffraction pattern at a distance d_0 behind an edge illuminated by a parallel beam.[7]

Figure 2 shows the calculated intensity variation $|u_P|^2$ as function of $\alpha f\cdot\rho/f$ for an electron wavelength 5.10^{-10}A ($\sim$60 kv) for perfect focus and for different distances d_0 between the edge and the focal plane; changing the sign of the defocusing merely changes the sign of the imaginary term in Eq. (9) and does not affect the intensity distribution in the pattern for the opaque edge. Furthermore, if the objective aperture is reduced from ρ_0/f to ρ_1/f and the defocusing separation is, simultaneously increased from d_0 to $d_1=(\rho_0/\rho_1)^2d_0$, the intensity pattern is simply expanded in the ratio

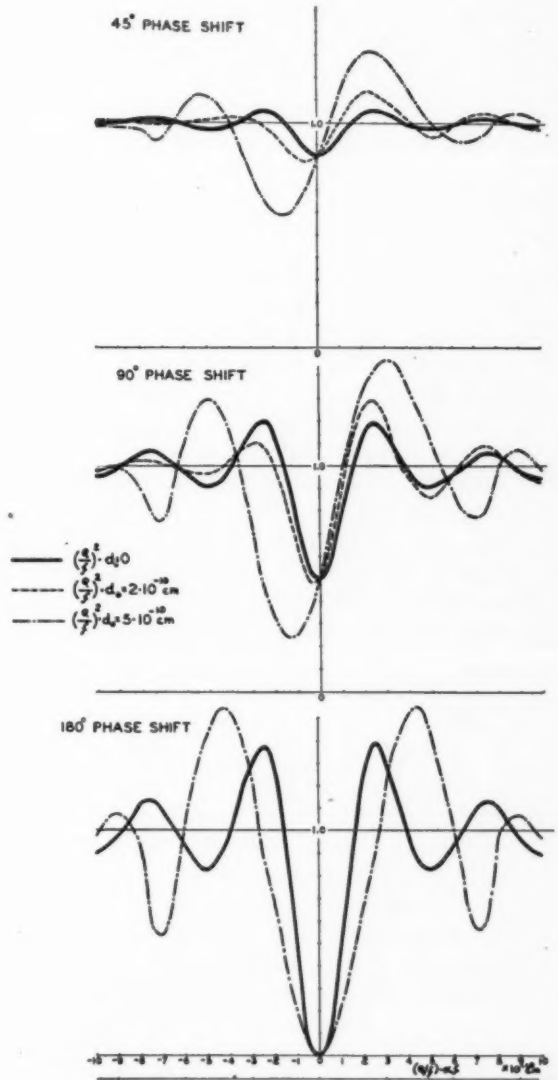

FIG. 3. The intensity distribution near the image of the edge of a transparent half-plane imparting a prescribed phase delay on the incident electron wave.

[5] See G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1922), p. 328.

[6] In the complete absence of an edge the intensity distribution near the center of the field is given, to the same order of approximation, by $1-[J_0(k\alpha\rho)]^2\exp[-ik\rho^2/(2f)]$.

[7] Equation (9) is not an adequate approximation under these circumstances.

ρ_0/ρ_1 without changing relative intensities at corresponding points. The curves in Fig. 2 indicate the transition from the diffraction pattern arising from the restricted lens opening to the simple Fresnel pattern of an edge. At large distances from the edge the intensity oscillations are attenuated most. This is to be expected since here the maxima and minima lie closest together and hence are most affected by the limited resolution of the system. The intensity patterns expected for infinite lens aperture are indicated for comparison.

II. PHASE CONTRAST AT THE EDGE OF A TRANSPARENT OBJECT

The phase contrast at the image of the edge of a transparent object film is obtained by adding the amplitudes for two complementary edge diffraction patterns, one of which is shifted in phase by the phase delay δ imparted by the transparent film:

$$\text{Intensity} = |u_P(\alpha) + e^{i\delta} u_P(-\alpha)|^2. \qquad (10)$$

The phase delay δ is here given by $\Delta k \cdot t = k \cdot t \cdot \Delta V/(2V)$, where ΔV is the inner potential of the film ($\sim$15 volts) and t is the film thickness. As mentioned above, a film 100A in thickness is seen to result in a phase shift of the order of $\pi/2$. Figure 3 shows the phase contrast for phase shift $\pi/4$, $\pi/2$, and π. The contrast is seen to increase with phase shift, reaching a maximum for π. For the last value, and generally at perfect forcus, the intensity distribution is symmetrical about the edge. Otherwise, changing the sign of the defocusing distance simply has the effect of changing the sign of α (i.e., the abscissas) on the intensity plot. One intensity distribution is the mirror image of the other through the edge.

The edge contrast is seen to increase, also, with the degree of defocusing. Even for perfect focus, however, the phase contrast at the edge may be considerably greater than the contrast arising from scattering beyond the effective objective aperture: For a 100A film of a substance with the effective atomic number 6 (carbon) and a specific gravity 1.5, the contrast arising from the scattering of 50-kv electrons beyond an aperture of 3.10^{-3} radian may be estimated to be approximately 5:4,[8] while the phase contrast (for 90° phase shift) is seen to be of the order of 5:2.

Caution is, of course, required in comparing the results here obtained with measured intensity distributions at the images of film edges. Qualitatively, the shift of the first minimum from the shadow to the clear side of the edge in passing from too strong to too weak objective currents is, in fact, observed.[9] Other features, such as the difference in contrast for the overfocused and underfocused image of the edge, which may be ascribed to the prismatic action of an edge of non-uniform thickness,[10] are, of course, not reproduced. The same applies to the contrast arising from scattering and absorption losses. Finally, most high resolution images of film edges have been obtained with large physical objective aperture, so that spherical aberration influences the intensity distribution. Even so, the calculations give some indication of the importance which phase contrast may assume in revealing specimen contours in normal electron microscope operation.

---

[8] See Zworykin, Morton, Ramberg, Hillier, and Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), p. 691.

[9] See, e.g., J. Hillier and E. G. Ramberg, J. App. Phys. **18**, 48–71 (1947), Fig. 10.

[10] See H. Boersch, Physik. Zeits. **44**, 32–38 (1943) and E. Kinder and A. Rocknagel, Optik **2**, 346–363 (1947).

Flow through a Rocket Nozzle with and without Vibrational Equilibrium

S. S. PENNER
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received November 8, 1948)

A procedure is outlined for calculating the maximum changes in the exhaust velocity of gases which can be introduced during flow through a rocket nozzle as the result of vibrational time lags. Representative calculations have been carried out for a number of propellant systems. These calculations indicate that the vibrational time lags cannot produce changes in exhaust velocity which exceed a few percent. The numerical values of the changes are found to depend not only on the nature of the gases but also on the occurrence or non-occurrence of chemical reaction during flow.

NOMENCLATURE

c = exhaust velocity if complete equilibrium is reached with respect to internal vibration.
c' = exhaust velocity if the vibrational heat capacity is frozen at the chamber temperature.
C_p = total heat capacity per mole at constant pressure including the heat capacity corresponding to dissociation.
C_p' = sum of molar rotational and translational heat capacities at constant pressure, including the heat capacity corresponding to dissociation.
$\bar{C}_p$ = average molar total heat capacity (including the heat capacity corresponding to dissociation) at constant pressure, between T_e and T_c.
$\bar{C}_p' = \bar{C}_p - \bar{C}_{vib}$ = average molar rotational and translational heat capacity (including the heat capacity corresponding to dissociation) at constant pressure, between T_e' and T_c.
C_{vib} = vibrational heat capacity per mole.
$\bar{C}_{vib}$ = average molar vibrational heat capacity between T_e and T_c.
$\bar{C}_{vib}'$ = average molar vibrational heat capacity between T_e' and T_c.
E = internal energy per mole.
H = enthalpy per mole.
$\Delta H_{T_e}^{T_c}$ = molar enthalpy change during flow through the rocket nozzle under conditions where complete equilibrium is reached with respect to internal vibration.
$\Delta H_{T_e'}^{T_c}$ = molar enthalpy change during flow through the rocket nozzle under conditions where the internal vibrations remain frozen at the chamber temperature.
p = pressure.
p_e = exit pressure.
p_c = chamber pressure.
R = gas constant per mole.
T = absolute temperature.
T_c = chamber temperature.
T_e = exhaust temperature if complete equilibrium is reached with respect to internal vibration.
T_e' = exhaust temperature if the vibrational heat capacity is frozen at the chamber temperature.
T_{vib} = vibrational temperature defined through the actual energy residing in the vibrational degrees of freedom.
u = linear flow velocity of the gases through a rocket nozzle, if the flow occurs parallel to the nozzle axis.
v = volume.

ADIABATIC expansion of the combustion products of a rocket motor through the nozzle[1] may occur so rapidly that there is not enough time to establish equilibrium with respect to the internal vibrations of the molecules. It is necessary to make a complete kinetic and flow analysis in order to determine the approach to flow with equilibrium partition of vibrational energy. The analysis of the flow process can be carried through, in a manner similar to that previously used to determine the presence or absence of equilibrium with respect to chemical reactions.[2] Some of the required kinetics data for calculations involving vibrational equilibrium may be found in the literature.[3–6]

Before proceeding with the laborious analysis of the history of vibrational energy distribution during flow, it is desirable to determine the effect of various vibrational distributions on the performance of a rocket motor. Accordingly, the maximum performance differences associated with the extremes of vibrational energy distribution which can be introduced during flow through the rocket nozzle have been estimated. Calculations have been carried out only for the case where complete equilibrium with respect to internal vibration is reached in the rocket chamber. Flow calculations with complete vibrational equilibrium and with internal vibrations frozen at the chamber temperature permit the determination of the maximum difference in exhaust velocity associated with the extremes of vibrational distribution introduced through zero or infinite vibrational time lag during expansion through the nozzle. The vibrational correction is assumed to be superimposed on the usual type of calculations for equilibrium flow or constant-composition flow.[1,2]

The method of calculation employed for the present study may be readily applied in those cases where the vibrational states are frozen at some temperature intermediate between the chamber and exhaust temperatures.

[1] F. J. Malina, J. Franklin Inst. **230**, 433 (1940).

[2] S. S. Penner and D. Altman, J. Franklin Inst. **245**, 421 (1948); D. Altman and S. S. Penner, J. Chem. Phys. **17**, 56 (1949).

[3] W. Jost, *Explosions- und Verbrennungsvorgänge in Gasen* (Verlag Julius Springer, Berlin, 1939).

[4] L. Landau and E. Teller, Physik. Zeits. Sowjetunion **1**, 34 (1936).

[5] A. Kantrowitz, J. Chem. Phys. **14**, 150 (1946).

[6] P. W. Huber and A. Kantrowitz, J. Chem. Phys. **15**, 275 (1947).

OUTLINE OF CALCULATIONS

Thermodynamic considerations lead to two alternate methods of calculation, both of which are described below.

A. Calculation of Exhaust Velocities from the Law of Conservative Energy

Flow through the rocket nozzle is assumed to be adiabatic. Conservation of energy leads to the relations

$$\tfrac{1}{2}c^2 = \text{constant } \Delta H_{T_e}{}^{T_c}, \tag{1}$$

$$\tfrac{1}{2}c'^2 = \text{constant } \Delta H_{T_e'}{}^{T_c}. \tag{2}$$

From Eqs. (1) and (2) it follows that

$$c/c' = (\Delta H_{T_e}{}^{T_c}/\Delta H_{T_e'}{}^{T_c}). \tag{3}$$

The exhaust temperatures T_e and T_e' are given by the relations[1, 2]

$$\begin{aligned}(p_e/p_c)^{R/\bar{C}_p} &= T_e/T_c,\\ (p_e/p_c)^{R/\bar{C}_p'} &= T_e'/T_c.\end{aligned} \tag{4}$$

Equations (3) and (4) permit the determination of the exhaust velocity ratio corresponding to flow with and without equilibrium with respect to internal vibration.

B. Calculation of Exhaust Velocities from the Enthalpy Theorem

The condition of conservation of energy may be expressed by the enthalpy theorem

$$H+\tfrac{1}{2}u^2 = E+pv+\tfrac{1}{2}u^2 = \text{constant},$$

or

$$\int_0^T C_p dT+\tfrac{1}{2}u^2 = \text{constant}. \tag{5}$$

Equation (5) holds for equilibrium partition of energy. If equilibrium is not maintained with respect to internal vibrations, then Eq. (5) should be replaced by the relation

$$\int_0^T C_p dT+\int_T^{T_{\text{vib}}} C_{\text{vib}} dT+\tfrac{1}{2}u^2 = \text{constant}. \tag{6}$$

A relation between c and c' may be obtained by applying Eqs. (5) and (6) to the two types of flow under consideration. The result is

$$\int_0^{T_e} C_p dT+\tfrac{1}{2}c^2 = \int_0^{T_e'} C_p dT + \int_{T_e'}^{T_c} C_{\text{vib}} dT+\tfrac{1}{2}c'^2, \tag{7}$$

or

$$c^2-c'^2 = 2\left[\int_{T_e'}^{T_c} C_{\text{vib}} dT+\int_{T_e}^{T_e'} C_p dT\right]. \tag{8}$$

Reference to Eqs. (3) and (8) shows that Eq. (3) determines the velocity ratio, whereas Eq. (8) permits calculation of the difference between c and c'. The two expressions together, therefore, permit calculation of the absolute values of both c and c'. This fact will not be particularly useful in practice, however, since the numerical value of c can be calculated in the usual manner[1] from the data required to determine C_p, C_p', $\Delta H_{T_e}{}^{T_c}$, and $\Delta H_{T_e'}{}^{T_c}$.

APPLICATION TO ROCKET MOTORS

Representative calculations of c and c' have been carried out and are described in this section for the hydrogen-fluorine and hydrogen-oxygen motors employing the reactants in stoichiometric proportions. Under these conditions it is justified to equate C_{vib}, as a first approximation, to the mean vibrational heat capacities of hydrogen fluoride and water, respectively.

The calculations are illustrated in detail for the hydrogen-fluorine rocket motor if chemical equilibrium exists during flow through the rocket nozzle. The required thermodynamic data were taken from unpublished work by M. Farber and are reproduced here: $T_c = 4915°\text{K}$; $T_e = 3640°\text{K}$; $c = 11344$ ft./sec. $= 3.46\times10^5$ cm/sec.; $\bar{C}_p = 19.82$ cal./°K/mole; C_{vib} for HF $= 2.23$ cal./°K/mole at 4915°K; C_{vib} for HF $= 1.90$ cal./°K/mole at 3640°K. From the preceding data it can be seen that $\bar{C}_{\text{vib}} \simeq 2.06$ cal./°K/mole and $\bar{C}_p' \simeq 19.82-2.06 = 17.76$ cal./°K/mole. From Eq. (4) it is found that $T_e' \simeq 3515°\text{K}$. Therefore $\Delta H_{T_e}{}^{T_c} \simeq 19.82\times(4915-3640) = 25{,}270$ cal./mole and $\Delta H_{T_e'}{}^{T_c} \simeq 17.76\times(4915-3515) = 24{,}864$ cal./mole. Substitution of the calculated enthalpy changes in Eq. (3) leads to the result $c/c' = 1.008$, or $c' = 11{,}253$ ft./sec. Therefore, the maximum difference between c and c' is seen to be less than one percent in the present case. Application of Eq. (8) instead of Eq. (3) is found to lead to the same result. Calculations for the hydrogen-fluorine motor have also been carried through for the case where flow through the nozzle occurs without composition change. In this case $\bar{C}_{\text{vib}}$ represents nearly 20 percent of $\bar{C}_p$. The ratio c/c' is now found to be increased to 1.034.

Calculations similar to those outlined for the hydrogen-fluorine rocket with chemical equilibrium during flow through the nozzle have been carried out for the hydrogen-oxygen system using unpublished calculations by R. B. Canright. The result is $c/c' = 1.006$. The conclusion may therefore be drawn that the state of vibrational energy distribution during flow through the nozzle is of no practical importance as regards the performance of the hydrogen-fluorine and hydrogen-oxygen motors if chemical equilibrium is maintained during flow. For both of the rocket motors considered, the exhaust velocity is lowered by less than one percent if the vibrational

states remain frozen at the chamber temperature. On the other hand, if constant-composition flow occurs through the rocket nozzle, then the exhaust velocity may be lowered by as much as 3.3 percent for the hydrogen-fluorine motor. On the basis of previous work[2,7] it appears probable that equilibrium is maintained during flow with respect to atomic reactions but that composition changes as the result of molecular reactions are less likely to occur. The actual correction to the exhaust velocity, corresponding to completely frozen vibrational energy states, may therefore be intermediate between the small values obtained for equilibrium flow and the larger values determined for constant-composition flow.

Calculations for other rocket motors indicate that the changes in exhaust velocity associated with frozen vibrational states are generally of the same order of magnitude as the results calculated for the hydrogen-fluorine rocket motor. For example, for the red fuming nitric acid-aniline rocket, at a weight ratio of nitric acid to aniline of 2.80, c/c' is found to have a value of 1.039 for equilibrium flow and 1.056 for constant-composition flow.

Reference to Eqs. (3) and (8) as well as to the calculated results of c/c' allows some qualitative conclusions regarding the effect of frozen vibrational energy states on the exhaust velocity. In general, c/c' will be large for systems in which $\bar{C}_{vib}$ constitutes a large fraction of $\bar{C}_p$. The numerical value of $\bar{C}_{vib}$ is nearly the same for flow with and without chemical equilibrium. On the other hand, $\bar{C}_p$ is always larger for equilibrium flow than for constant-composition flow. It would therefore be expected that c/c' is larger for constant-composition flow than for equilibrium flow. This conclusion is borne out by the numerical results given earlier in this discussion.

The ratio c/c' depends on the chamber temperature, chamber pressure, and exhaust pressure. This can be shown by writing Eq. (3) in the form

$$c/c' = \{\bar{C}_p(T_c - T_e)/\bar{C}_p' T_c[1-(p_e/p_c)^{R/\bar{C}_p'}]\}^{\frac{1}{2}}.$$

The vibrational heat capacity $\bar{C}_{vib}$ is relatively large for gaseous mixtures containing polyatomic molecules. Thus the red fuming nitric acid-aniline system contains CO_2 and H_2O with sizable vibrational heat capacities. As the result, c/c' is found to be greater for the red fuming nitric acid-aniline rocket than for the hydrogen-fluorine rocket. In conclusion, it should be pointed out that the heavy polyatomic molecules such as CO_2, which have large vibrational heat capacities, may also be the molecules with relatively long relaxation times which are most likely to lead to flow with frozen vibrational energy states.[5,6] In other words, in some of the systems for which c/c' is largest the exhaust velocity is likely to decrease from c to c'. Just how large this decrease actually is can be determined only after a complete study of the flow process and of the relaxation times in gaseous mixtures has been made.

[7] S. S. Penner, J. Am. Chem. Soc. **71**, 788 (1949).

A Note on the Substitution Principle for Steady Gas Flow

R. C. PRIM III
Naval Ordnance Laboratory, White Oak, Maryland and Princteon University, Princeton, New Jersey
(Received November 8, 1948)

The Substitution Principle established by Munk and Prim for flows of a gas having constant specific heats is extended to the wider class of fluids having a state equation of the form $\rho=P(p)S(s)$. The equations of motion for this class of fluids is reduced to a canonical form involving only the reduced velocity vector and the pressure. It is shown that explicit elimination of the pressure from these equations is possible only for those fluids whose state equation has the form $\rho=p^kS(s)$ with $k<1$. Relations between the Mach number and stagnation pressure and the reduced velocity field are discussed for the class of fluids under investigation.

INTRODUCTION

THE following Substitution Principle was established[1] by Munk and Prim for steady flow in the absence of extraneous force fields of a non-viscous, thermally non-conducting gas with constant specific heats:

If in any flow field the densities are multiplied by m and the velocities by $1/(m)^{\frac{1}{2}}$, where m is a positive function which is a constant along each individual streamline, the new flow field will be a possible flow of the same gas having the same pressure distribution and Mach number distribution as the original flow.

This Substitution Principle is particularly useful in clarifying the study of rotational gas flows—that is, flows in which the entropy and/or the total flow energy vary from streamline to streamline. For example, the function m can always be so chosen that the total flow energy of the derived (or "substitute" flow) will be uniform throughout the flow field. Or, alternatively, in a region between two successive shock fronts an isentropic flow can be substituted for a non-isentropic one. Hence questions involving flows which are both non-isoenergetic and non-isentropic can be resolved by the study of isoenergetic flows alone.

Furthermore, by the choice $m=a^2$, where a is the ultimate velocity magnitude along a streamline, we obtain the canonical formulation[2,3] of the equations of gas dynamics in terms of the reduced velocity vector $\bar{w}(\bar{w}\equiv\bar{v}/a)$ and the pressure alone:

$$\bar{w}\cdot\operatorname{grad}\bar{w}+(\gamma-1/2\gamma)(1-w^2)\operatorname{grad}\ln p=0 \quad (1)$$

and

$$\operatorname{div}[(1-w^2)^{1/\gamma-1}\bar{w}]=0, \quad (2)$$

where p denotes the pressure and γ the (constant) ratio of specific heats. The condition of integrability for the pressure field is derived from (1) in terms of the $\bar{w}$ field alone:

$$\operatorname{curl}\left[\frac{\bar{w}\times\operatorname{curl}\bar{w}}{1-w^2}\right]=0. \quad (3)$$

The set of Eqs. (2) and (3) then represent necessary and sufficient conditions involving the $\bar{w}$ field alone. They have proved particularly useful for the purposes of discovering families of exact solutions and establishing general geometric and kinematic properties of rotational flow fields.[4–8]

The purpose of the present note is to determine to what extent these results can be extended to flows of fluids having a more general state equation than that of a perfect gas

$$\rho=Kp^{1/\gamma}\exp(-s/c_p). \quad (4)$$

(ρ denotes the density, s the specific entropy, and c_p the (constant) specific heat at constant pressure. K is a dimensional constant.) In particular, it will be determined (a) for what equations of state $\rho=\rho(p,s)$ the Substitution Principle is valid and a consequent canonical formulation of the equations of flow in terms of the pressure and reduced velocity fields alone can be obtained; and (b) for which of these equations of state the pressure can be eliminated from these canonical equations.

THE SUBSTITUTION PRINCIPLE

The equations governing the steady flow in absence of extraneous force fields of a compressible fluid without viscosity or thermal conductivity can be written in the form

$$\operatorname{div}\rho\bar{v}=0 \quad (5)$$

[1] M. Munk and R. Prim, "On the multiplicity of steady gas flows having the same streamline pattern," Proc. Nat. Acad. Sci. **33**, 137–141 (1947).
[2] M. Munk and R. Prim, "On the Canonical form of the equations of steady motion of a perfect gas," Naval Ordnance Laboratory Memorandum 9169 (June 1947). J. App. Phys. **19**, 957 (1948).
[3] B. Hicks, P. Guenther, and R. Wasserman, "New formulations of the equations for compressible flow," Quar. App. Math. **V**, 357–361 (1947).
[4] P. Nemény and R. Prim, "Some geometric properties of plane gas flow," J. of Math. and Phys. **XXVII**, 130–135 (1948).
[5] R. Prim, "On the existence of steady gas flow in plane isothermal streamline patterns," Bull. Am. Math. Soc. **54**, 1009–1012 (1948).
[6] P. Nemény and R. Prim, "Some Properties of Rotational Flow of a Perfect Gas," Proc. Nat. Acad. Sci. **34**, 119–124 (1948).
[7] R. Prim, "Extension of Crocco's theorems to flows having non-uniform stagnation enthalpy," Phys. Rev. **73**, 186 (1948).
[8] R. Prim, "On a family of rotational gas flows," Quar. App. Math. **VI**, 319–325 (1948).

$$\bar{v}\cdot\text{grad}\bar{v}+(1/\rho)\,\text{grad}p=0 \tag{6}$$

$$\bar{v}\cdot\text{grad}s=0 \tag{7}$$

$$\rho=\rho(p,\,s). \tag{8}$$

In order to establish the Substitution Principle we shall introduce the "kinetic flux" vector $\bar{q}$ defined by

$$\bar{q}=(\rho)^{\frac{1}{2}}\bar{v}. \tag{9}$$

In terms of this vector, (5) and (6) can be written

$$\text{div}\bar{q}+\tfrac{1}{2}(\bar{q}\cdot\text{grad}\,\ln\rho)=0 \tag{10}$$

and

$$\bar{q}\cdot\text{grad}\bar{q}-\tfrac{1}{2}(\bar{q}\cdot\text{grad}\,\ln\rho)\bar{q}+\text{grad}p=0. \tag{11}$$

Also, since from (8)

$$\text{grad}\,\ln\rho=\left(\frac{\partial\,\ln\rho}{\partial p}\right)_s\text{grad}p+\left(\frac{\partial\,\ln\rho}{\partial s}\right)_p\text{grad}s, \tag{12}$$

(7) can be written

$$(\bar{q}\cdot\text{grad}\,\ln\rho)-\left(\frac{\partial\,\ln\rho}{\partial p}\right)_s\bar{q}\cdot\text{grad}p=0. \tag{13}$$

Equations (10), (11), and (13) are clearly invariant under any change of dependent variables leaving $\bar{q}$, p, and $(\bar{q}\cdot\text{grad}\,\ln\rho)$ unchanged *provided that* $(\partial\,\ln\rho/\partial p)_s$ *is a function of* p *alone*. This permitted change of dependent variables is of the form

$$p^*=p;\quad \rho^*=m\rho;\quad \bar{v}^*=(1/(m)^{\frac{1}{2}})\bar{v}, \tag{14}$$

where m is a constant along each individual streamline ($\bar{v}\cdot\text{grad}m=0$). The provision that $(\partial\,\ln\rho/\partial p)_s$ be a function of p only limits the validity of this substitution to fluids having a "product" equation of state; that is, an equation of state (8) of the form

$$\rho=P(p)S(s). \tag{15}$$

The invariance of the Mach number under the substitution (14) follows from the identity

$$\left(\frac{\partial\,\ln\rho}{\partial p}\right)_s=\frac{(\partial\rho/\partial p)_s v^2}{\rho v^2}=\frac{M^2}{q^2} \tag{16}$$

for fluids having a product equation of state.

The validity of this substitution is not affected by the presence of shock discontinuities in the flow. The familiar equations relating the flow quantities on the two sides (1 and 2) of a shock front can be written in the form

$$(\rho_1)^{\frac{1}{2}}q_1\sin\alpha_1=(\rho_2)^{\frac{1}{2}}q_2\sin\alpha_2$$

$$p_1+q_1^2\sin^2\alpha_1=p_2+q_2^2\sin^2\alpha_2$$

$$\frac{q_1^2}{\rho_1}+2\int_0^{p_1}\frac{dp}{\rho(p,\,s_1)}=\frac{q_2^2}{\rho_2}+2\int_0^{p_2}\frac{dp}{\rho(p,\,s_2)}$$

(α denoting the angle between flow direction and shock) in which it is immediately evident that the substitution (14) is permissible.

Hence, the Substitution Principle stated in the Introduction for flows of a perfect gas is valid in the same form for flows of any fluid having an equation of state of the form $\rho=P(p)S(s)$.

The Principle is also valid for flows of a non-homogeneous incompressible fluid. (In place of (13) we have $\bar{q}\cdot\text{grad}\,\ln\rho=0$. The Mach number is then of course zero throughout.)

THE CANONICAL EQUATIONS

From Eq. (6) we have the Bernoulli or energy equation

$$v^2+2h(p,\,s)=a^2=2h(p_0,\,s) \tag{17}$$

wherein the enthalpy h is defined by

$$h(p,\,s)=\int_0^p\frac{dp}{\rho(p,\,s)}, \tag{18}$$

and the ultimate velocity a and the stagnation pressure p_0 are constants along each individual streamline in a region free from shocks. (The quantities a and $h(p_0,\,s)$ are unchanged by a shock front; but p_0 and s are changed discontinuously.)

We now set about the elimination of the redundancy of dependent variables which the Substitution Principle reveals by introducing the reduced velocity vector $\bar{w}$ through

$$\bar{w}\equiv\bar{v}/a \tag{19}$$

(equivalent to the choice $m=a^2$).

The dynamic Eq. (6) then becomes

$$\bar{w}\cdot\text{grad}\bar{w}+\frac{1}{\rho a^2}\text{grad}p=0 \tag{20}$$

and (17)

$$h(p,\,s)=\tfrac{1}{2}a^2(1-w^2)=h(p_0,\,s)(1-w^2). \tag{21}$$

By the use of (21), (20) can be written

$$\bar{w}\cdot\text{grad}\bar{w}+\tfrac{1}{2}(1-w^2)\frac{\text{grad}p}{\rho(p,\,s)h(p,\,s)}=0. \tag{22}$$

Now *for the product equation of state* (15) the term

$$\rho(p,\,s)h(p,\,s)=P(p)\int_0^p\frac{dp}{P(p)} \tag{23}$$

is a function of p alone. (It can further be shown that *only* for the product equation of state is this the case.) Hence, letting

$$H(p)\equiv\int_0^p\frac{dp}{P(p)};\quad P(p)=\frac{1}{[dH(p)/dp]} \tag{24}$$

we can now write (22) in the canonical form

$$\bar{w}\cdot\text{grad}\bar{w}+\tfrac{1}{2}(1-w^2)\,\text{grad}\,\ln H(p)=0 \tag{25}$$

involving $\bar{w}$ and p alone. Also (21) becomes

$$H(p)=H(p_0)(1-w^2), \tag{26}$$

and from (25) and (26) we obtain the important relation

$$\text{grad}\ \ln H(p_0)=2\frac{\bar{w}\times\text{curl}\bar{w}}{1-w^2}, \tag{27}$$

connecting the $\bar{w}$ and p_0 fields. The integrability condition for (25) is the same as that (3) for a perfect gas.

The introduction of the reduced velocity vector into the continuity equation (5) yields

$$\text{div}\rho\bar{w}=0$$

which for the product equation of state takes on the canonical form

$$\text{div}P(p)\bar{v}=0. \tag{28}$$

In (25) and (28), then, we have a formulation of the equations (Eqs. (5)–(8)) of gas dynamics in terms of $\bar{w}$ and p alone for any fluid with a product equation of state. (For a non-homogeneous incompressible fluid, (10) and (11) immediately yield such a formulation in terms of $\bar{q}$ and p). Given a p field and a $\bar{w}$ field satisfying (25) and (28), the stagnation pressure p_0 is determined by (26). The ultimate velocity a can then be assigned arbitrarily for each streamline to obtain a velocity field $\bar{v}$. The entropy is then determined by

$$S(s)=[2H(p_0)/a^2], \tag{17a}$$

and the density by (15). As will be shown later, the Mach number is given by (34). For a shock-free region, the entropy s may be assigned on each streamline instead of the ultimate velocity a, and the latter computed from (17a).

ELIMINATION OF THE PRESSURE FROM THE CONTINUITY EQUATION

We have shown that any product equation of state leads to an integrability condition (3) in terms of $\bar{w}$ only. We now ask which product equations of state permit the elimination of $P(p)$ from the continuity equation (28) (by the use of (25)) to obtain a form of this equation involving $\bar{w}$ only (such as (2) for perfect gases).

To answer this question we rewrite (28) in the form

$$\text{div}\bar{w}+\frac{d\ \ln P(p)}{d\ \ln H(p)}\bar{w}\cdot\text{grad}\ \ln H(p)=0. \tag{29}$$

The factor $(\bar{w}\cdot\text{grad}\ \ln H(p))$ can be expressed in terms of $\bar{w}$ by the use of (25). If the resulting form of (29) is not to involve the pressure, it is then necessary that

$$[d\ \ln P(p)/d\ \ln H(p)]=\text{constant}.$$

It can be readily shown from (24) that this implies that the function $P(p)$ is of the form

$$P(p)=Cp^k, \quad (k<1), \tag{30}$$

where C and k are constants.

Hence, the most general product equation of state permitting the elimination of the pressure from the equations is

$$\rho=p^kS(s). \quad (k<1). \tag{31}$$

The equations obtained are the same as those ((1) and (2)) for a perfect gas with γ set equal to $1/k$. Given a $\bar{w}$ field satisfying (2) and (3) (with γ replaced by $1/k$), a consistent pressure field can be obtained by integration from (1). The other flow quantities can then be determined as described above for the more general product equation of state. It will be observed that for the exponential state equation (31), the pressure field can be multiplied by any positive constant without violating the differential equations, but that this is not the case for the general product equation of state (15).

It should be noted that the variety of physically meaningful state equations represented by (31) is not substantially greater than the family (4) of state equations for perfect gases with $1<\gamma\leq 5/3$, because such considerations as the positiveness of temperature and of internal energy preclude the existence of substances with state equations of the form (31) with $k<0$, $S(s)<0$, or $S'(s)>0$.

RELATION BETWEEN MACH NUMBER AND REDUCED VELOCITY

It is well known that for a perfect gas the Mach number and the magnitude of the reduced velocity vector are substantially equivalent quantities connected by the relation

$$M^2=(2/\gamma-1)(w^2/1-w^2). \tag{32}$$

For the general equation of state $\rho=\rho(p,s)$ the Mach number can by use of (19) and (21) be expressed as follows

$$\begin{aligned} M^2&=[\partial\rho(p,s)/\partial p]v^2\\ &=2[\partial\rho(p,s)/\partial p]h(p,s)(w^2/1-w^2)]\\ &=M^2(w^2,p,s). \end{aligned} \tag{33}$$

For the product equation of state, (33) can be written

$$\begin{aligned} M^2&=2[d\ \ln P(p)/d\ \ln H(p)](w^2/1-w^2)\\ &=M^2(w^2,p), \end{aligned} \tag{34}$$

and we see that only for the power law state equation (31) does (33) take on the form

$$M^2=M^2(w^2).$$

(For this case

$$M^2=[2k/1-k][w^2/1-w^2].) \tag{35}$$

Stabilization of Breakdown Voltage and Measurement of Some Critical Currents in a Glow Tube

A. B. PAL
Department of Electrical Engineering, Princeton University, Princeton, New Jersey
(Received January 14, 1949)

This paper describes a special technique by which it is possible to stabilize the breakdown voltage of a glow tube having zirconium electrodes in pure argon at a pressure of 8.5 mm Hg. A brief account is also given of some experimental methods by which it is possible to find the order of magnitude of two critical currents, whose concepts have been used in the theory of spark discharge at low pressures. One of these critical currents is the "threshold current," which is defined as the maximum steady-state current in a glow tube just before a spark. The other critical current is the maximum electronic current (attained during the development of a spark) at cathode before space-charge distortion becomes appreciable. The methods described below were confined to two tubes having zirconium cathodes of area 0.3 sq. cm, the electrode spacing being 3.5 mm. One of these two tubes had a zirconium anode, and the other one had a thorium anode. The tubes contained pure argon at a pressure of 8.5 mm Hg. For both the tubes, the order of the "threshold current" was found to be 10^{-8} amp., and the order of the other critical current (electronic current at cathode) was estimated as 10^{-3} amp.

STABILIZATION OF BREAKDOWN VOLTAGE

IN order to produce a spark discharge in a glow tube, the presence of some initial ionization current i_0 is necessary. Experiment shows that the "time lag" may be several seconds, if the initial ionization current $i_0 \leqslant 10^{-14}$ amp. If the value of i_0 is fixed and kept much above the value 10^{-14} amp., then the breakdown voltage will have a definite value, and the value of the time lag will be almost zero. It is found that a 100-w lamp placed near the tube with a zirconium cathode produces an initial ionization current whose value is of the order of 10^{-11} amp., and so the breakdown voltage assumes a definite value. We may define the breakdown voltage as the value of applied voltage which produces a spark discharge almost immediately, when the initial ionization current has a given value, say, 10^{-11} amp. In defining the breakdown voltage, the arbitrary value of i_0 must be stated, as the value of the breakdown voltage depends upon the value of i_0.

The value of the breakdown voltage E_b at low pressures of the order of a few mm. Hg is mainly governed by the following three factors:

(1) Nature and purity of the enclosed gas.
(2) Nature of the cathode surface.
(3) Initial ionization current i_0.

If these three factors are kept unchanged, the value of the breakdown voltage E_b will also remain unchanged. In actual practice, it is very difficult to keep the first two factors unchanged. A heavy discharge may produce impurities and also change the nature of the cathode surface, resulting in a different value of E_b for the next spark. Continuous glowing of the cathode is also undesirable, as it may heat the cathode and so change its work function. Such a heating may lower the breakdown voltage by several volts. Thus, in an actual test, a tube with zirconium cathode had an E_b equal to 157 volts at room temperature. On glowing the cathode with a d.c. current of about 50 ma for a few minutes, the cathode became red hot, and the value of E_b fell to 112 volts, i.e., the fall in E_b was 45 volts. On cooling the tube to room temperature, the value of E_b was restored to the original value 157 volts.

In order to control the first two factors, we must select a suitable cathode material and a suitable gas for filling the tube. A suitable material for the cathode is pure zirconium, and a suitable gas is pure argon. Impurities must be removed by repeated induction heating of the electrodes, while the tube is connected to a pump. The tube must be baked for a long time. There must be long traps between the tube and the pump cooled by liquid air, in order to prevent mercury vapor and other impurities from entering the tube. When the tube is ready, its cathode surface requires further cleaning in order to make the surface stable for a long time. This requires intermittent bombardment of the cathode surface with positive ions. The bombardment must occur in almost instantaneous blows, repeated after sufficiently long intervals, so that heat produced by bombardment radiates away during the vacant intervals and does not penetrate into the body of the cathode. The surface cleaning may last for several hours before it is satisfactorily complete. The writer got very good results by continuing this process for as long a time as 100 hours! The breakdown voltage then becomes constant and no change is detected in its value even after two months.

Figure 1 shows a very efficient circuit for cleaning the cathode surface. It enables us to control the discharge current, keep its value almost constant during a discharge, control the duration of a discharge, and also control the interval between two successive discharges. This circuit thus meets all

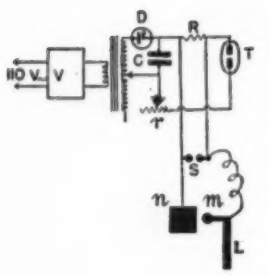

FIG. 1. Circuit for cathode surface cleaning (D=diode, V=variac, C=100 μf, R=0.1 megohm, r=1000 ohms).

the requirements of the cleaning process. Its action may be explained as follows:

The condenser C acquires a d.c. voltage (about 500 volts) much above the breakdown value E_b for the tube T and so a permanent glow current flows through the tube. When S is open, the glow current is very small due to the presence of the resistor R, which has a resistance equal to 0.1 megohm. When S is closed, the value of the glow current rises to a high value, which is controlled by the tapping point on the resistor r, which has a value of 1000 ohms. The gap S is mechanically opened and closed by the vibrating lever L (attached to an ordinary electric bell) and the rugged terminals n and m. The period of vibration of the lever L controls the interval between two successive heavy discharges. The duration of a heavy discharge is controlled by adjusting the position of n, which changes the time of contact between n and m. It can be shown that the value of the heavy discharge current during the time of contact between n and m remains approximately constant, if the value of C is about 100 μf and the time of contact is about 1/1000 second.

When the cathode surface becomes clean, the glow on its surface shows uniform brightness at all points. Unclean surfaces give glows whose brightness varies from point to point on the cathode surface. Thus, the appearance of the glow also helps detect any impurities in the surface of the cathode.

MEASUREMENT OF THE FIRST CRITICAL CURRENT

The concept of this critical current (also called the "threshold current") has been used by Loeb.[1] We shall denote this current by i_e (Loeb denotes it by $\bar{i}$). The writer has measured the threshold current i_e by several methods, and all these methods agree in giving the order of magnitude of this critical current as 10^{-8} amp. for the two tubes used by the writer. We shall briefly describe these methods one by one.

First Method

Suppose i_p is initial ionization current produced by some light illuminating the zirconium cathode. There is also some current i_r due to initial ionization produced by stray radiations, such as cosmic rays. The value of i_r changes at random. The total current due to initial ionization, i.e., i_0, is thus given by

$$i_0 = i_p + i_r. \tag{1}$$

Experiment shows that i_0 has large random fluctuations, if $i_p \leqslant 10^{-16}$ amp. These fluctuations are due to the existence of the random component i_r. As we raise the value of i_p, the percent random variation in i_0 decreases. The random changes in i_0 produce random changes in E_b. Experiment shows that if we illuminate the zirconium cathode by direct sunlight at noon, then the value of i_p is of the order of 10^{-9} amp. With such a high value of i_p, the percent random fluctuation in i_0 becomes almost nil, and the value of E_b remains constant. The value of time lag for such a large value of i_0 is also almost nil.

Now, consider the circuit shown in Fig. 2, in which T is a tube having a zirconium cathode illuminated by direct sunlight at noon, and I is a known current, which can be set at any desired value in the range 10^{-9} to 3×10^{-6} amp. The current I can be produced by a device using a photo-cell exposed to light of controlled intensity.

The current I charges the condenser C. Let the voltage built up in C be denoted by E. When E is well below the breakdown value E_b, then the tube current $i \ll I$, and the rise of E with time t is almost linear. As $E \rightarrow E_b$, the tube current approaches the critical value i_e, which we want to measure. Suppose $I \gg i_e$. As the tube current i (before breakdown) can rise only to the maximum value i_e, the charging current $(I-i)$ remains almost equal to I, since $I \gg i_e$. The voltage E thus rises almost linearly with time t from E_i (extinction voltage) to E_b, and we get a series of discharges, which are heard as a series of clicks in the sounder. The interval T (between two successive clicks) is given by

$$T = C(E_b - E_i)/I, \quad \text{since } I \gg i_e. \tag{2}$$

As we lower the value of I, so that ultimately $I \rightarrow i_e$, then (2) cannot hold, since the value of the charging current $(I-i)$ drops to a negligible value as $i \rightarrow i_e$ near the breakdown condition. Suppose E rises to E_0 (which is slightly below E_b) in a time T_0, the value E_0 being such that for $E \leqslant E_0$, the tube current i remains negligibly small. Let the time taken by E to rise from E_0 to E_b be t_d. The total time T (between two consecutive clicks) is thus given by

$$T = T_0 + t_d = C(E_0 - E_i)/I + t_d. \tag{3}$$

If $I \gg i_e$, then $T_0 \gg t_d$, and $T \approx T_0$. As $I \rightarrow i_e$, the value of $t_d \rightarrow \infty$. Thus, if we gradually lower the value of I from a large initial value, and notice the value of I when T becomes very large, then this value of I will be roughly equal to i_e. In this way,

[1] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley & Sons, Inc., New York, 1939), pp. 420–423.

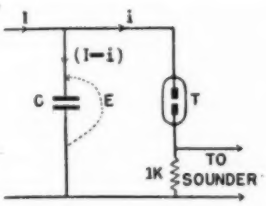

FIG. 2. Circuit for measuring i_e by methods 1 and 2.

it is possible to find an approximate value of i_e. Experiment shows that the interval between two clicks becomes very large for values of I around 2×10^{-8} amp. This shows that the order of magnitude of the current i_e is 10^{-8} amp.

Second Method

If the zirconium cathode of a glow tube be illuminated by a lamp of 100W instead of direct sunlight at noon, we notice appreciable random fluctuations in i_0. As the value of E_b depends upon the value of i_0, there will also be some noticeable fluctuations in E_b. In this case, the value of t_d in Eq. (3) will depend upon chance. In other words, the value of t_d will be erratic. For values of $I \gg i_e$, the value of $t_d \ll T_0$, and so the total interval T will be approximately constant. But if we lower the value of I, so that I remains only slightly above the critical current i_e, then the value of the erratic component t_d will be very large, and this will make the total interval T also erratic in value. Measurements of the interval t_d for different values of the capacity C have shown that for values of I around 2×10^{-8} amp. the interval t_d becomes very large and erratic. Hence, the value of i_e is roughly equal to 2×10^{-8} amp., and its order of magnitude is definitely 10^{-8} amp.

Third Method

The principle used in this method will be clear from the circuit shown in Fig. 3.

Suppose the d.c. supply voltage E_0 is kept several volts above the breakdown value E_b of the tube T. When S_2 is closed and S_1 is open, a glow current passes through the tube T. The value of this glow current depends upon the resistance R. If we open the switch S_2, the glow current will charge the condenser C. In this way, a voltage is built up in C and as the value of this voltage rises, the value of the glow current falls, and ultimately the current drops to a value which is almost zero. Suppose the voltage on C rises from 0 to x volts in a time t. The value of x can be measured as follows. After a time t, when the voltage on C acquires the value x, we suddenly lower E_0 to 0. If we again raise the value of E_0, we shall get a spark discharge in T (heard as a loud click in the sounder) for some value E_b' of E_0, such that

$$E_b' = E_b + x \quad \text{or} \quad x = E_b' - E_b.$$

As we know E_b' and E_b, we can at once find the value of x. We can measure x for different values of t, and then plot x *vs.* t. From this plot we can find the value of dx/dt, and hence of the tube current (Cdx/dt), for any value of t. As t increases, the value of the tube current falls, and ultimately drops to a value which is almost zero. As the value of the tube current for tube voltage nearly equal to E_b is i_e, and the tube current drops to almost zero for a slight fall in value from E_b, the minimum measurable value of the tube current in the plot of tube current *vs.* time must be of the order of i_e. The writer measured x for several values of t, and plotted x *vs.* t for different values of R, such as 22, 44, 66, 88, and 110 megohms. The value of C was 0.5 μf. From these plots, the values of tube current were computed for different values of t. These values showed that the tube current had a minimum value of the order of 10^{-8} amp. before suddenly dropping to a value which was almost zero. Hence, the order of magnitude of the critical current i_e is also 10^{-8} amp.

Fourth Method

For this method, we use the circuit arrangement shown in Fig. 4.

Here R is a very high resistance—several thousand megohms! The writer used the wooden spacing between two terminals fixed on a wooden block as a high resistance, and found its value quite suitable. The value of R changes with time, but remains pretty constant for the short interval during which we use R in our experiment. C is simply the stray capacity. As we raise the supply voltage above E_b, the tube may be in the breakdown condition. But the tube current is kept low by the high resistance R, and unless it rises to the critical value i_e, a breakdown cannot occur. We can increase the value of tube current, however, by raising the supply voltage. So we increase the supply voltage to a value E_e, when a spark discharge is produced and is heard as a click in the sounder. The energy is supplied by the capacity C, which is discharged. Just before the breakdown, the voltage across the tube is nearly equal to E_b, and the tube current is of order i_e. Hence, we have the simple relation:

$$i_e = (E_e - E_b)/R. \qquad (4)$$

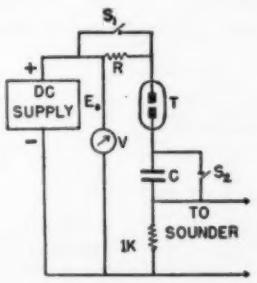

FIG. 3. Circuit for measuring i_e by method 3 (V = voltmeter, R = 22, 44, 66, 88, or 110 megohms, C = 0.5 μf).

In an actual experiment, the following values were observed for a tube.

$$E_c = 263 \text{ volts}, \quad E_b = 154 \text{ volts}.$$

Putting these values in Eq. (4), we get

$$i_e = 109/R. \tag{5}$$

To find R, we set $C = 5000\ \mu\mu f$, and keep the supply voltage equal to E_c. We note the time T taken by C to build up its voltage from 0 to E_b, when a discharge will pass through the tube. The following values were observed for T. $T = 17.2$, 17.0, 16.8, 17.0, 17.1. Mean value of $T = 17.0$ sec. These values clearly indicate that R remains pretty constant for several minutes, and its value is quite reliable for our purpose. Now, we use the relation:

$$E = E_c(1 - e^{-t/RC}).$$

At $t = T$, E becomes equal to E_b. Thus

$$E_b = E_c(1 - e^{-T/RC}).$$

In this equation, we know the values of all the quantities, except R. We can, therefore, compute the value of R. Calculation shows that $R = 3.8 \times 10^9$ ohms. Putting this value of R in Eq. (5), we get $i_e = 2.9 \times 10^{-8}$ amp. The order of magnitude of i_e is therefore definitely 10^{-8} amp.

Fifth Method

Consider the circuit shown in Fig. 5. As before, I is a current which can be set at any value in the range 10^{-9} to 3×10^{-6} amp. by means of a suitable device. When the switch S is closed, the voltage across C_2 remains 0, and the tube breaks down each time the voltage built up in C_1 exceeds the value E_b for the tube T. We thus get a series of discharges, which produce clicks in the sounder at regular intervals. The interval between two successive clicks depends upon the values of I, C_1, E_b, and E_i (the extinction voltage). If $I \gg i_e$, then this interval is $C_1(E_b - E_i)/I$.

What happens to the discharges, if S is opened? Experiment shows that if S is opened just after a discharge, the next discharge may not occur under certain conditions for values of I below a certain critical value I_0.

Let us see how I_0 is related to i_e, the threshold current. As the voltage across T reaches E_b, the tube current rapidly builds up to i_e, just before the spark. Thus, when the voltage across T is a little below E_b, the value of the tube current is an appreciable fraction of i_e. The tube current indicated in Fig. 5 by i_2 is thus of the same order as i_e. Hence, the determination of i_2 gives a value whose order is the same as that of i_e, in which we are interested.

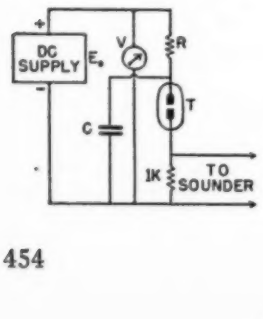

FIG. 4. Circuit for measuring i_e by method 4.

Now, let us see how it is possible to measure i_2. As the resistance in series with the tube T is $1K$ (or 1000 ohms), the voltage drop across it by a current of the order of 10^{-8} amp. is almost nil, and so we shall ignore it. Let the instantaneous value of the voltage built up in C_1 be E_1, and that in C_2 be E_2. The voltage acting across the tube is then $(E_1 - E_2)$. If the circuit components are so chosen that (for $i_2 \leqslant i_e$)

$$dE_1/dt = dE_2/dt, \tag{6}$$

then the value of $(E_1 - E_2)$ will remain constant. Notice dE_2/dt is almost nil, when $(E_1 - E_2)$ is well below the breakdown value E_b, since the tube current i_2 becomes appreciable (and of the order of i_e) only when the applied voltage $(E_1 - E_2)$ becomes close to the breakdown value E_b. Hence, condition (6) is satisfied only when i_2 is of the order of i_e. When this condition is satisfied, then the tube voltage $(E_1 - E_2)$ remains permanently below E_b, and so we cannot get another discharge.

The value of i_2 can be determined in terms of I_0, C_1, and C_2, if condition (6) is satisfied. It is obvious that

$$i_1 = C_1 dE_1/dt, \quad \text{and} \quad i_2 = C_2 dE_2/dt.$$

Thus, by Eq. (6), we get

$$i_1/C_1 = i_2/C_2. \tag{7}$$

But $I_0 = i_1 + i_2$, and so we get the value of i_2 given by

$$i_2 = I_0 C_2/(C_1 + C_2). \tag{8}$$

The writer performed an experiment to measure i_2 by Eq. (8). Several values of C_1 and C_2 were tried. The value of I was gradually reduced till on opening S the next discharge was delayed by a long time. The value of I, when the discharge time becomes very large, is then very nearly equal to the critical value I_0 in Eq. (8). (As mentioned before, we open S just after a discharge, which is one of the series of discharges which appear when S is kept closed.) In performing this experiment, the zirconium cathode must be illuminated by light, as i_2 cannot flow continuously in dark. The writer used a lamp of 100w for this purpose. In dark, i_0 (current due to initial ionization) fluctuates at random, and this makes i_2 also unsteady, so that Eq. (7) cannot be satisfied. Thus, with the tube having a thorium anode in dark, no critical value of I could be observed. On illuminating the zirconium cathode with a lamp of 100w, however, it

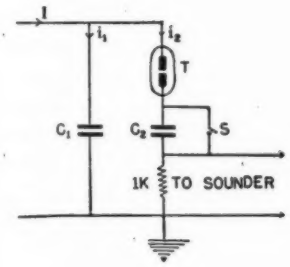

FIG. 5. Circuit for measuring i_s by method 5.

was possible to detect and measure the critical value I_0.

As a result of several observations, the values of i_2 computed by using Eq. (8) were seen to be of the order of 10^{-8} amp.

It will be seen from the above discussion that all the above five methods agree in giving the order of magnitude of the threshold current i_s equal to 10^{-8} amp. for the two tubes used by the writer. These tubes had zirconium cathodes of surface area 0.3 sq. cm, the anode-cathode gap being about 0.4 cm, and the enclosed gas was argon at a pressure of 8.5 mm Hg. The anode of one of these tubes was of zirconium, and that of the other one of thorium. The electrodes were properly seasoned so that each tube had a definite and stabilized breakdown voltage.

MEASUREMENT OF THE SECOND CRITICAL CURRENT

The second critical current, which we shall denote by i_s, was introduced by Schade[2] in his theory of the formative time lag. According to this theory, the main part of the formative time lag is spent in raising the cathode electronic current from its initial value i_0 to the critical value i_s. The space-charge distortion is assumed to be serious only when the cathode current exceeds the critical value i_s. According to Schade's theory, the formative time lag T_f is given by the equation:

$$T_f=(t_i/\epsilon)\ \ln[1+\epsilon(i_s/i_0)], \qquad (9)$$

where t_i means the time taken by the positive ions to cross the anode-cathode gap, which is usually of the order of one centimeter. The value of t_i is of the order of 10^{-6} sec. The quantity ϵ is a function of the over-voltage $(E-E_t)$, where E is the applied voltage, and E_t is the threshold voltage. As $E\rightarrow E_t$, the quantity $\epsilon\rightarrow 0$. Hence, in case E is set at the threshold value, Eq. (9) reduces to the simple form

$$T_f=(t_i i_s/i_0), \quad \text{or} \quad i_s=(T_f i_0/t_i). \qquad (10)$$

Schade verified Eq. (9) in a rough manner. His experiments can give no value for the critical current i_s. Schade regards the critical current i_s, however, as the minimum glow current, and assumes its value to be of the order of 10^{-5} amp. According to the writer's estimate, however, the critical current i_s seems to be of the order of 10^{-8} amp. This may be shown as follows. As each electron starting from the cathode produces several electrons on its way to the anode, the current i_s assumes a higher value I_s at the anode. We may set

$$I_s=i_s p, \qquad (11)$$

where p denotes a multiplying factor. Theoretically, it is hard to find the value of the factor p. It is possible, however, to estimate roughly the factor p from some experimental values of currents at the cathode and at the anode which are not very different in order of magnitude from the currents i_s and I_s. Experiment shows that the value of i_0 (initial ionization current) for the tube with zirconium electrodes exposed to direct sunlight at noon is of the order of 10^{-9} amp. When the applied voltage is just below E_b, the current at anode has the value of order 10^{-8} amp., as shown by the above five methods of measuring this current. If we assume the positive ions are not liberating appreciable number of electrons from the cathode, then the order of p is seen to be 10, since the electronic current at cathode is 10^{-9} amp., and the electronic current at anode is 10^{-8} amp. As a rough approximation, we may set p in Eq. (11) equal to 10, since the space-charge distortion is assumed to be negligible upto the value i_s. In order to compute i_s from Eq. (11), we still need the value of the second unknown current I_s, which is the minimum glow current flowing in the tube just after a spark discharge gets started. The value of this current may be measured experimentally by the circuit shown in Fig. 6. First, we close the switch S, and lower E_2 to 0 volt. Then slowly increase the voltage E_1, so that voltage built up in C is nearly the same as E_1, the value of the time constant Cr being very small. When $E_1=E_b$, we get a spark discharge in the tube T. Then we lower E_1 to 0, in order to restore the tube to normal condition, and then raise E_1 to a value slightly below the breakdown value E_b. Now, open S, and gradually raise the voltage E_2. As (E_1+E_2) exceeds the value E_b, the tube breaks

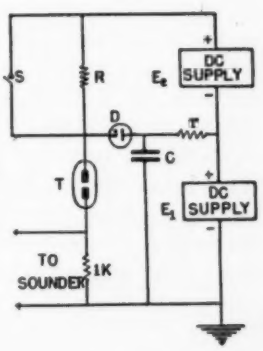

FIG. 6. Circuit for measuring minimum glow current (E_1=voltage just below the breakdown value, $Cr\ll 1$ second, D=diode, R=66, or 110 megohms).

[2] L. B. Loeb and J. M. Meek, *The Mechanism of the Electric Spark* (Stanford University Press, California, 1940), pp. 12–20.

down. But the tube current (a weak glow discharge) is controlled by the high resistance R. As we increase E_2 still further, the glow current is increased. With the rise of the glow current, the voltage across the tube falls below E_1, the diode circuit is at once closed, the condenser C sends a heavy discharge through the tube T, and this discharge is heard as a loud click in the sounder. At this moment, the voltage across T is roughly equal to E_b, and as all the current flowing in the tube (just before the condenser C discharges) also flows through the resistance R, we have the simple relation:

$$\text{Glow current just after breakdown} = (E_1+E_2-E_b)/R. \quad (12)$$

The current given by Eq. (12) is expected to be quite small, since the voltage across T is a little less than E_b, and the glow current becomes appreciably high only when the voltage across T has an appreciable fall below E_b. As the exact value of the voltage across T just before the moment of hearing the loud click is not known, the value given by this method is only an approximation to the lowest glow current I_s, which can possibly flow just after a breakdown. The following values refer to an actual experiment:

(1) $R=66$ megohms, $E_b=225$ volts, $(E_1+E_2)=260$ volts.

Putting these values in Eq. (12), we get $I_s=5.3\times10^{-7}$ amp.

(2) $R=110$ megohms, $E_b=225$ volts, $(E_1+E_2)=280$ volts.

Putting these values in Eq. (12), we get $I_s=5.0\times10^{-7}$ amp. This experiment shows that the order of magnitude of the minimum glow current I_s in Eq. (11) may be roughly taken as 10^{-7} amp. As the order of the factor p in Eq. (11) has been shown to be 10, we see that the order of magnitude of the critical current i_s is 10^{-8} amp.

It will be noted that the value of i_s assumed by Schade, namely, 10^{-5} amp., differs very widely from the writer's estimate which is 10^{-8} amp. In order to decide which estimate is the correct one, we may resort to experiment and measure the order of i_s by using Eq. (10), which is based on Schade's theory. For this purpose, we have to measure the value of T_f and i_0. The value of t_i may be taken as 10^{-6} sec., so far as the order of magnitude is concerned. The writer produced i_0 by illuminating the zirconium cathode of a tube with a lamp of power 15W, placed at a distance of about four inches from the cathode. In order to control the intensity of light (and hence the value of i_0) given by the lamp, the voltage across the lamp terminals was supplied through a variac, having 130 scale divisions. When the variac was set at 0 division, the cathode of the tube was in perfect darkness. In this case, the value of T_f was several hundred secs. and also highly erratic, since the only source of i_0 was stray radiations which change at random. On gradually increasing the intensity of light, the value of T_f was found to fall, and its different values for a given intensity of light (and hence a given value of i_0) became less erratic. With the variac set at 70 divisions, the value of T_f became of the order of 50 sec. When the variac was set at its maximum position, the value of T_f reduced to about 2 sec. The value of i_0 (corresponding to the maximum position of the variac) was still too small to be measured by a direct and convenient method. On reducing the distance of lamp from the cathode to 1/4 of its previous value, the value of the intensity of light falling on the cathode was increased 16 times, and the new value of i_0 was thus made 16 times greater than its previous value. This is, however, an approximation, as the lamp was not a point source of light. Now, the value of i_0 was 5×10^{-13} amp., which was determined by using the "second method" described in an article already published by the writer.[3] The value of i_0 for the previous position of the lamp may, therefore, be taken as $(1/16)\times(5\times10^{-13})$ or 3×10^{-14} amp. approximately. The corresponding value of T_f was 2 sec., as already stated. We are, now, in a position to compute the order of magnitude of i_s by using Eq. (10). Taking only the orders of magnitude of the quantities T_f, i_0, and t_i, we get at once the order of magnitude of i_s equal to 10^{-8} amp., which agrees with the writer's estimate. Hence, the order of magnitude of the second critical current i_s is 10^{-8} amp., and not 10^{-5} amp., as assumed by Schade. The validity of Schade's estimate could not be checked by the rough manner in which Schade tested his theory.

It will be noticed that the currents mentioned in this paper are total currents, and not current densities. As the electrode area is 0.3 sq. cm, the order of the current density is not very different from that of the total current.

ACKNOWLEDGMENTS

I wish to acknowledge my considerable debt to Dr. H. C. Rentschler under whose guidance the above work was done, and also I sincerely appreciate the kind help given by the EE Department, Princeton University. The writer is especially indebted to Professor C. H. Willis, Chairman of the EE Department, and Professor W. C. Johnson of the same department, for their valuable comments and helpful discussions.

[3] A. B. Pal, Rev. Sci. Inst. 19, 531 (1948).

High Frequency Total Emission Loading in Diodes*

NICHOLAS A. BEGOVICH
Hughes Aircraft Company, Culver City, California
(Received August 16, 1948)

An expression is derived for the total emission conductance produced by the electrons returned to the cathode of a diode operating with a retarding off-cathode field. For the retarding voltage and diode spacing used, the assumption of a linear potential distribution is justified. Examples of the variation of the total emission conductance with retarding voltage and frequency are illustrated.

I. INTRODUCTION

AN interesting high frequency loading phenomenon occurring in very close-spaced diodes is reported by a number of investigators.[1-3] The experimental set-up is briefly as follows. The "Q" of the cathode-anode cavity of the diode is measured as a function of the d.c. retarding voltage on the anode.[4] The frequency used for the measurements is 3×10^9 c.p.s.[5] At a very large d.c. retarding voltage, the "Q" of the cavity is the same whether the cathode is heated or not. With the cathode heated, the "Q" is measured as the retarding voltage is decreased in magnitude. The "Q" sharply decreases and reaches a minimum value before any measurable value of d.c. current is collected at the anode (i.e., before any appreciable number of electrons reach the anode).

Clearly, the loading of the cavity must be produced by the electrons that are returned to the cathode by the retarding anode field. On an average, because of the finite to and fro transit time, the electrons returning to the cathode must have an energy greater than their emission energy. This extra energy is supplied by the high frequency field and is converted into heat when the returning electrons strike the cathode. This loading has been named "total emission loading" since it is produced by the total emission of the cathode.[1,6]

In the normal operation of vacuum tubes as amplifiers, the anode current is always only a small fraction of the total emission current. Consequently, there exists in front of the cathode a potential minimum and a resulting retarding off-cathode field. This retarding field drives most of the electrons (~90 percent) back into the cathode. If these electrons return to the cathode with an energy greater than their emission energy, the input circuit of the tube will require a greater input driving power for a given power output.

II. THEORY

For the very close cathode-anode spacing used in these experiments and with retarding voltages of the order of 1 or 2 volts, the potential distribution between the cathode and anode is very close to linear. This assumption is suggested by the experimental evidence of maximum loading before an appreciable number of electrons strike the anode of the diode. An additional justification for this assumption will be given in Section III. In the equations that follow, c.g.s.-practical units will be used.

Figure 1 shows the configuration of the diode consisting of two parallel infinite planes a (the cathode) and b (the anode). The spatial dependence for all quantities will be given by the x variable measured from the cathode, i.e., the problem is assumed one-dimensional. To calculate the energy of the returning electrons, the force equation must be integrated.[7]

$$m\ddot{x}=-10^7eE$$
$$=-10^7(e/d)(V_{b0}-V_{b1}\cos\{\omega t+\phi\}). \quad (1)$$

The stationary (d.c.) retarding voltage at the b plane is V_{b0}; the time dependent voltage is V_{b1}. Note that the sign of the electron charge is written into the equation so that when e appears it will always mean the absolute value of the electron charge. At the cathode ($x=0$), the electron's initial condition is $v=v_0$ at time $t=0$. Integrating (1) and using the cited initial condition, the velocity at time t is

$$\dot{x}=v_0-(l/d)\left(V_{b0}t+\frac{V_{b1}}{\omega}\{\sin\phi-\sin[\omega t+\phi]\}\right), \quad (2)$$

where $l=10^7e/m$. The position at time t is

$$x=v_0t-(l/d)\left(\tfrac{1}{2}V_{b0}t^2+\frac{V_{b1}}{\omega^2}\times\{\omega t\sin\phi+\cos(\omega t+\phi)-\cos\phi\}\right). \quad (3)$$

* The results of this calculation were reported at Pasadena, June 24–26, 1948, meeting of the American Physical Society.

[1] C. N. Smyth, Nature **157**, 841 (1946).

[2] A. van der Ziel and A. Versnel, Nature **159**, 640 (1947).

[3] D. R. Hamilton, J. K. Knipp, and J. B. H. Kuper, *Klystrons and Microwave Triodes* (McGraw-Hill Book Company, Inc., New York, 1948), p. 156.

[4] This experiment was also performed on the input cavity of a triode. (See reference 3.)

[5] An exception to this is the van der Ziel measurements which were made at 4×10^7 and 5.2×10^7 c.p.s.

[6] A. van der Ziel and A. Versnel, Philips Research Reports **3**, 13 (1948).

[7] This method of calculating the transit time has been used by a number of writers. Compare C. J. Bakker and G. deVries, Physica **1**, 1045 (1934); and K. S. Knol, M. J. O. Strutt, and A. van der Ziel, Physica **5**, 325 (1938).

FIG. 1. Basic diode configuration.

The stationary transit time t_0 (transit time when $V_{b1}=0$) for the to and fro path of the electron is given by one solution of (3) with x set equal to zero, the other solution being the trivial one $t_0=0$.

$$t_0=2v_0d/lV_{b0}. \tag{4}$$

With V_{b1} different from zero, the transit time T can be written as

$$T=t_0+p\delta_1+p^2\delta_2+\cdots, \tag{5}$$

where $p=V_{b1}/V_{b0}$. Clearly, p is the proper parameter for the transit time expansion since the ratio of V_{b1} to V_{b0} will have a direct influence on the perturbation of t_0. Using (4) and (5) in (3) with x set equal to zero, expanding the resultant expression in powers of p, and separately setting the coefficient of each one equal to zero, give

$$\delta_1=\tfrac{1}{2}(t_0/\theta^2)\{(\cos2\theta-1)\cos\phi-(2\theta-\sin2\theta)\sin\phi\} \tag{6}$$

and

$$\delta_2=-(\delta_1^2/t_0)+(\delta_1/\theta)\{(\sin2\theta\cos\phi-(1-\cos2\theta)\sin\phi\}, \tag{7}$$

where $\theta=\omega t_0/2$ (the half-way transit angle).

The velocity v_1 of the returning electron at the cathode will be given by substituting (4) to (7) in (2). Squaring the resultant expression, and multiplying by $m/2$, will give the difference in the emitted and returning kinetic energy for an electron leaving the cathode with a velocity v_0 at a phase angle ϕ for the potential V_{b1}. The average energy interchange per electron for electrons leaving the cathode with a velocity v_0 will be the integration of this expression over a complete period of the phase angle ϕ.

$$\overline{\tfrac{1}{2}m(v_1^2-v_0^2)}=m(v_0p/\theta)^2(1-\theta\sin2\theta-\cos2\theta). \tag{8}$$

The number of electrons per second per cm^2 $(\Delta n/\Delta t)$ leaving the cathode in the velocity range v_0 to v_0+dv_0 is given by the Maxwell-Boltzmann expression

$$(\Delta n/\Delta t)=(I_0/leV_T)\exp(-v_0^2/2lV_T)v_0dv_0, \tag{9}$$

where I_0 is the emission current amps./cm^2, and $V_T=kT_c/10^{+7}e$ k being Boltzmann's constant 1.380×10^{-16} ergs/deg., and T_c being the absolute cathode temperature. The energy interchange per second per cm^2 between the cathode and the emitted and returned electrons in the velocity range v_0 to v_0+dv_0 is given by multiplying (8) by (9). Dividing this product by 10^7, the energy interchanged for the velocity packet dv_0 is

$$\Delta P=(I_0/V_T)(v_0p/l\theta)^2(1-\theta\sin2\theta-\cos2\theta)\times\exp(-v_0^2/2lV_T)v_0dv_0\ \frac{\text{watts}}{\text{cm}^2}. \tag{10}$$

Since the transit angle θ is a function of v_0, it will be more convenient to write (10) completely in terms of θ.

Introducing the following parameters

$$h=V_{b0}^2l(2\omega^2d^2V_T)^{-1}=1.021\cdot10^{19}(\omega d)^{-2}T_c^{-1}V_{b0}^2, \tag{11}$$

and

$$\alpha=\omega d(lV_{b0})^{-1}, \tag{12}$$

(10) can be written as

$$\Delta P=2I_0p^2h(l\alpha^2)^{-1}(1-\theta\sin2\theta-\cos2\theta)e^{-h\theta^2}\theta d\theta. \tag{13}$$

The critical emission velocity, the velocity of emission which separates the electrons reaching the anode from those returning to the cathode, is

$$v_l^2=2l|V_{b0}|, \tag{14}$$

or in terms of the transit angle

$$\theta_l=\alpha v_l=2.115\cdot10^{-7}fd(V_{b0})^{-\frac{1}{2}}, \tag{15}$$

where f is the frequency of V_{b1}. When V_{b1} is different from zero, (14) should be modified. However, as will be seen shortly this will be an unnecessary correction. Using (15) for the upper limit in the integration of (10) (see Appendix A), and defining a total emission stream conductance G_s as

$$P=\tfrac{1}{2}V_{b1}^2G_s \text{ watts/cm}^2, \tag{16}$$

give

$$G_s=\frac{4I_0kT_c}{10^7eV_{b0}^2}\Big[e^{-V_{b0}/V_T}\{(h+1)\cos2\theta_l+h(\theta_l\sin2\theta_l-1)\}-1+\frac{h+2}{h}\times e^{-1/h}{}_1F_1(1/2;3/2;1/h)\Big]\text{ mhos/cm}^2, \tag{17}$$

where ${}_1F_1$ is the confluent hypergeometric function. A curve is given on page 279 for ${}_1F_1(1/2;3/2;1/h)$ in Jahnke and Emde,[8] which in their notation is

[8] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943).

denoted as $M(1/2; 3/2; 1/h)$. Since

$$ {}_1F_1(1/2; 3/2; 1/h) = h^{\frac{1}{2}} \int_0^{1/\sqrt{h}} e^{t^2} dt, \qquad (18) $$

the table given on page 32 of Jahnke and Emde for the exponential integral can be used since it is more convenient for numerical work. For the examples to be considered, the exponential term in (17) can be completely neglected since, to justify the constant field assumption, V_{b0}/V_T will always be greater than 10. Consequently,

$$ G_s \cong (4I_0kT_c/10^7 ehV_{b0}^2)\{(h+2)e^{-1/h} \times {}_1F_1(1/2; 3/2; 1/h) - h\} \text{ mhos/cm}^2, \qquad (19) $$

where h is given by (11). For a very low frequency $V_{b1}\cos\omega t$ ($\theta \sim 0$), the first term in the bracket expression in (19) is 4/3 (see Appendix B), so

$$ G_s \underset{\omega \to 0}{\cong} (32I_0m\omega^2d^2k^2T_c^2)/(3e^3V_{b0}^4 10^{21}) $$

$$ = (32mI_0\omega^2d^2V_T^2)/(3e10^7V_{b0}^4) \text{ mhos/cm}^2. \qquad (20) $$

The frequency square dependence of the total emission conductance has been experimentally obtained by van der Ziel.[2] For an extremely high frequency $V_{b1}\cos\omega t$ (see Appendix C),

$$ G_s \underset{\omega \to \infty}{\cong} (2e10^7 I_0)/(m\omega^2 d^2) \text{ mhos/cm}^2. \qquad (21) $$

III. EXAMPLES

(1) To illustrate (17), the constants given by Kuper[3] for the frequency and diode spacing will be used. $f = 3\times10^9$ c.p.s.; $d = 8.89\times10^{-3}$ cm; $T_c = 10^3$ deg.; $I_0 = 0.1$ amps./cm^2. Since $V_{b0}/V_T = 11.605\,V_{b0}$ for the above given value of the cathode temperature, the exponential term in (17) can be neglected and (19) can be used for the calculation. To justify the assumed field distribution, the stationary potential distribution was calculated for the above given values of d, T_c, and I_0. The results are shown in Fig. 2. For retarding potentials greater in magnitude than 1 or 2 volts, the assumption of a linear potential distribution is certainly justified. In Fig. 3, the total emission conductance and its reciprocal, the resistance, is plotted as a function of the retarding voltage V_{b0}. The rapid increase in loading as the retarding voltage is decreased is clearly exhibited. The resistance curve is in good agreement with that given by Kuper,[3] who cites the experimental results obtained at the M.I.T. Radiation Laboratory by Dr. H. V. Neher, when the shunt resistance of the cavity used in Dr. Neher's experiment is considered. Since the stationary anode current is related by the Boltzmann expression

$$ I_a = I_0 \exp(-|V_{b0}|/V_T) \qquad (22) $$

to the emission current, the value of the anode current for the range of V_{b0} shown in Figs. 3 and 4 is of the order of 10^{-6} amps./cm^2 or less. Consequently, the electrons that cross to the anode will have a completely negligible effect on the conductance.

(2) Another interesting example is the variation of loading with frequency for a constant value of retarding voltage. The values of d, T_c, and I_0 used, were those of example 1. Figure 4 shows the variation of the total emission conduction as a function of frequency for retarding voltages of 3, 5, and 7 volts. Kuper[3] states that the loading should decrease as the frequency increases, though experimental results seem to indicate the opposite. As Fig. 4 shows, the behavior at 3×10^9 c.p.s. is an increase in loading as the frequency is increased. The decay of the loading for very high frequencies is also very interesting and can be explained as follows. For a very high frequency $V_{b1}\cos\omega t$, and consequently large transit angles, the energy interchange between the field and electrons decreases (note Eq. (8)). In addition, a small change in the emission velocity will produce a large change in the transit angle and a greater averaging out among the different electron velocity groups. The asymptotic value of the conductance is given by (21). In the derivation of (21) the assumptions used in Appendix A must be noted. Namely, though (21) appears to be independent of V_{b0}, it was assumed that $V_{b0} \gg V_T$ so that the first term in (17) could be neglected.

(3) Another interesting calculation is on the distance the electrons, that produce the maximum loading, penetrate in the retarding field. This can be obtained by differentiating (13) with respect to θ and setting the resultant equal to zero. The follow-

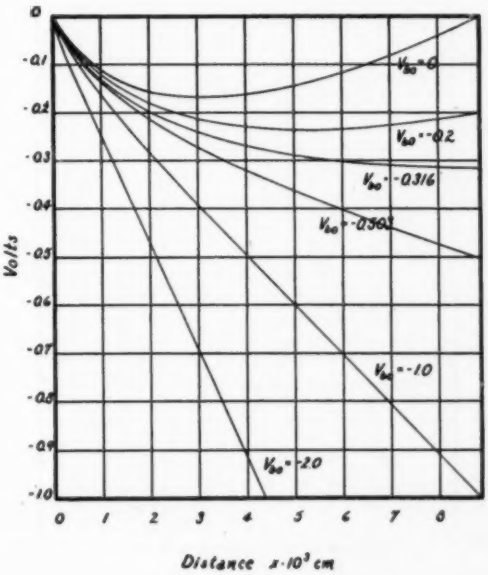

FIG. 2. Stationary potential distribution for different values of anode retarding voltage.

ing transcendental equation is obtained for the halfway transit angle (θ_M) for the electrons producing maximum loading,

$$\cos 2\theta_M\{1+2\theta_M^2(1-h)\} = 1+2h\theta_M^2(\theta_M \sin 2\theta_M - 1). \quad (23)$$

Using the data given in example (1), θ_M was calculated and is shown in Fig. 5 plotted against the retarding voltage V_{b0}. In addition, the ratio of the maximum distance (x_M) the electrons penetrate in the field, to the diode spacing d, is plotted.

CONCLUSIONS

The large magnitude of loading calculated above suggests that, for close spaced high frequency tubes, the cathode-first grid region should not be in the radio frequency circuit. Though the above calculation was made assuming a linear potential distribution, while a triode in normal operation would have a potential distribution different from linear outside the cathode (see Fig. 2), the order of magnitude of the loading obtained will be very nearly the same. This can be clearly seen from the behavior of the experimental curves.[2,3] An obvious, though mechanically difficult, way of excluding the potential minimum region from the radio frequency circuit is to operate a tetrode as a triode with the first and second grids as the input cavity region. By proper spacing the potential minimum can be made to reside in the cathode-first grid region, thus obtaining the benefit of shot noise reduction by space charge without undesirable loading.

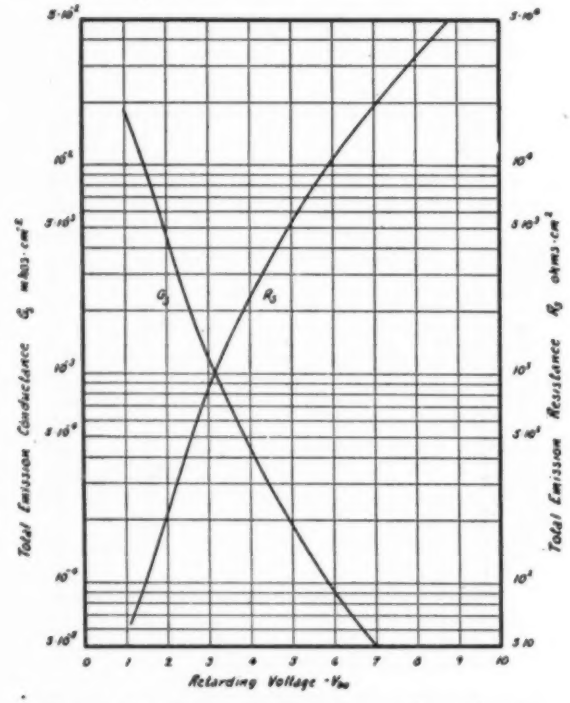

FIG. 3. G_s and R_s *versus* anode retarding voltage.

APPENDIX A

By repeated integration by parts, the integral in (13) becomes

$$\int_0^{\theta_1} e^{-h\theta^2}(1-\cos 2\theta-\theta\sin\theta)\theta d\theta = \frac{1}{2h^2}\left\{e^{-h\theta_1^2}[(h+1)\cos 2\theta_1+h(\theta_1\sin 2\theta_1-1)] -1+(h+2)\int_0^{\theta_1}\sin 2\theta e^{-h\theta^2}d\theta\right\}. \quad (a)$$

Because of the rapid decay of the exponential in the $\sin 2\theta$, integral, the upper limit can be replaced by infinity. The maximum possible error introduced by this change of the upper limit can be evaluated by setting $\sin 2\theta$ equal to one and integrating from θ_1 to infinity. The value of this integral is given in terms of the error function as

$$\int_{\theta_1}^{\infty} e^{-h\theta^2}d\theta = \tfrac{1}{2}\left(\frac{\pi}{h}\right)^{\frac{1}{2}}\left\{1-\operatorname{erf}\left(\frac{V_{b0}}{V_T}\right)^{\frac{1}{2}}\right\}. \quad (b)$$

The $\exp(-h\theta^2)\sin 2\theta$ integral with infinity as the upper limit can be evaluated in the following manner. Let

$$u=\int_0^{\infty} e^{-hx^2}\sin 2px dx. \quad (c)$$

Differentiating with respect to the parameter p gives

$$du/dp = 2\int_0^{\infty} xe^{-hx^2}\cos 2px dx. \quad (d)$$

Using integration by parts, the integral can be written in terms of u as

$$\frac{du}{dp}=\frac{1}{h}(1-2pu). \quad (e)$$

The solution of this first order differential equation is

$$u=\frac{1}{h}e^{-p^2/h}\int e^{p^2/h}dp, \quad (f)$$

where the constant of integration is zero since $u=0$, when $p=0$. Changing the integration variable in (f) and setting

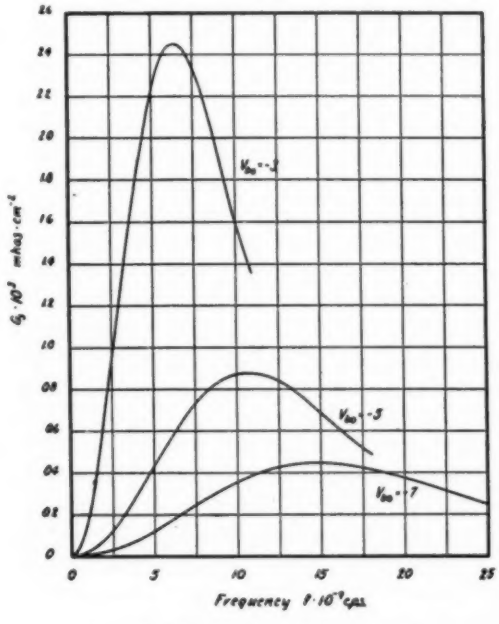

FIG. 4. G_s *versus* frequency of V_{b1}.

$p=1$, give

$$\int_0^\infty e^{-hx^2}\sin 2x\,dx=\frac{1}{h^{\frac{1}{2}}}e^{-1/h}\int_0^{1/h^{\frac{1}{2}}}e^{t^2}dt. \qquad \text{(g)}$$

Using the expansion for e^{t^2} and integrating term by term,

$$h^{\frac{1}{2}}\int_0^{1/h^{\frac{1}{2}}}e^{t^2}dt=\sum_{m=0}^{\infty}\frac{\Gamma(2m+1)}{\Gamma(2m+2)\Gamma(m+1)}\left(\frac{1}{h}\right)^m. \qquad \text{(h)}$$

By the use of Legendre's duplication formula[9] and the definition of the confluent hypergeometric function,[10]

$$h^{\frac{1}{2}}\int_0^{1/h^{\frac{1}{2}}}e^{t^2}dt={}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{1}{h}\right). \qquad \text{(i)}$$

Substituting (i) in (g) and the resulting expression in (a) gives (17).

APPENDIX B

To obtain the small transit angle expression for (19), the ${}_1F_1$ and $e^{-1/h}$ can be written in series form and multiplied together. However, an easier method is to make use of Kummer's first transformation.[11]

$$e^{-1/h}{}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{1}{h}\right)={}_1F_1\left(1;\frac{3}{2};-\frac{1}{h}\right). \qquad \text{(j)}$$

Using the expansion shown in (h)

$${}_1F_1\left(1;\frac{3}{2};-\frac{1}{h}\right)=1-\frac{2}{3h}+\frac{4}{15h^2}-\frac{8}{105h^3}+\frac{16}{945h^4}-\cdots. \qquad \text{(k)}$$

[9] E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford University Press, New York, 1935), p. 225.
[10] See reference 9, p. 260.
[11] See reference 9, p. 261.

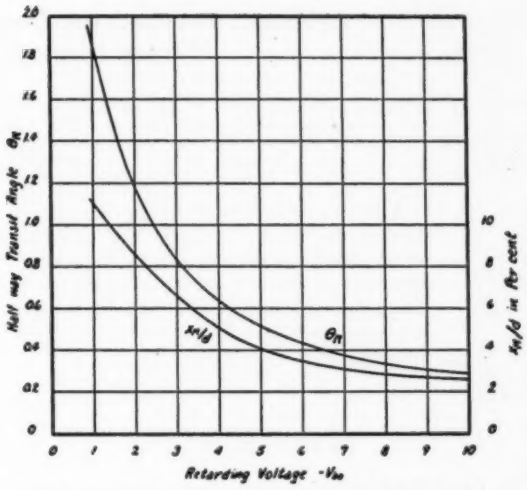

FIG. 5. θ_M and distance ratio *versus* V_{b0} for the electron velocity packet producing the maximum loading.

Substituting (k) in the bracket expression in (19)

$$\left\{(h+2)e^{-1/h}{}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{1}{h}\right)-h\right\}=\frac{4}{3}\left[1-\frac{4}{5h}+\frac{12}{35h^2}-\frac{32}{315h^3}+\cdots\right]. \qquad \text{(l)}$$

APPENDIX C

Using the asymptotic expansion for ${}_1F_1(1/2;\,3/2;\,1/h)$ for large $(1/h)$ gives

$${}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{1}{h}\right)\cong\frac{1}{2}he^{1/h}\left[1+\frac{h}{2}+\frac{3}{4}h^2+\frac{15}{8}h^3+\cdots\right]. \qquad \text{(m)}$$

Substituting (m) in the bracket expression in (19),

$$\left\{(h+2)e^{-1/h}{}_1F_1\left(\frac{1}{2};\frac{3}{2};\frac{1}{h}\right)-h\right\}\cong h^2\left[1+h+\frac{9}{4}h^2+\cdots\right]. \qquad \text{(n)}$$

A Visual Method for Demonstrating the Path of Ultrasonic Waves through Thin Plates of Material

CHARLES J. BURTON* AND R. BOWLING BARNES*
Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut
(Received December 1, 1948)

A visual method for studying the path of ultrasonic waves through thin plates of material has been devised, and a set of photographs showing the complexity of ultrasonic reflection and transmission by metal plates has been obtained. An attempt has also been made to estimate the critical angles of total reflection for dilatation and shear waves by visual examination of the reflection of a supersonic beam. Using these data, Young's modulus and the shear modulus for a $\frac{1}{16}$-inch aluminum sheet have been determined and the values are compared with those obtained using a previously described electrical method and with published data.

INTRODUCTION

ULTRASONIC techniques for determining the mechanical properties of metals, plastics, and glasses have been the subject of considerable study during recent years. Details concerning various phases of this work have been described in a number of papers.[1-8] Schneider and Burton[7] have shown that a rotating plate technique in which ultrasonic transmission is plotted electrically as a function of the angle of incidence of the waves allows determination of the velocities of dilatation and shear waves in the plate. From these data, Poisson's ratio and the mechanical moduli may be determined. Elastic constants of materials measured in this way are in good agreement with previously published values. Another paper[8] described an improved optical technique for visualizing ultrasonic waves, the method having been developed in an effort to clarify some of the data which appeared in the earlier work.

It is the purpose of the present paper to describe briefly a series of experiments showing the nature of the reflection and transmission of an ultrasonic beam incident upon an aluminum plate at various angles of incidence. It will be shown that the critical angles of total reflection for longitudinal and shear waves can be rapidly determined by inspection of the optical pattern and that these data can be used to obtain with fair accuracy the mechanical moduli and Poisson's ratio. Although not capable of high precision in its present stage of development, the method does have the advantage that the mechanical properties may be determined very rapidly from a single set of measurements.

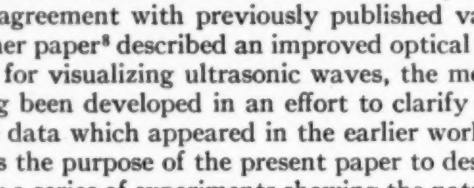

FIG. 1. Schematic diagram of optical arrangement used for making sound waves visible.

EXPERIMENTAL

The apparatus used in the present set of experiments is the same as that described in detail in a previous paper.[8] It comprises an optical system of the schlieren type which makes possible the direct visualization of an acoustic wave in a liquid medium as a result of refractive index gradients which are set up by passage of the compressional wave through the medium. An X-cut quartz crystal is employed as the electromechanical transducer, providing ultrasonic vibrations of 5-megacycle frequency, or at odd harmonics thereof. An electronic oscillator capable of supplying a high frequency potential of 300 volts is used to drive the quartz crystal. A device for supporting thin plates of vari-

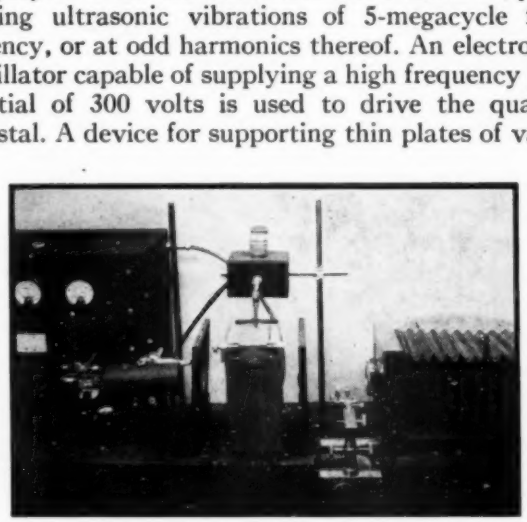

FIG. 2. Photograph of the experimental equipment for making sound waves visible: oscillator and point light source at left, ultrasonic tank in center and camera at right.

* Present Address: American Optical Company, Southbridge, Massachusetts.

[1] L. Bergmann, *Ultrasonics* (John Wiley & Sons, Inc., New York, 1938).
[2] W. P. Mason and T. G. Kinsley, Bell Lab. Record (May 18, 1948).
[3] J. W. Ballou and S. Silverman, J. Acous. Soc. Am. 16, 113 (1944).
[4] D. L. Arenberg, J. Acous. Soc. Am. 20, 1 (1948).
[5] A. W. Nolle, J. App. Phys. 19, 753 (1948).
[6] S. Siegel, J. Acous. Soc. Am. 16, 26 (1944).
[7] W. C. Schneider and C. J. Burton, J. App. Phys. 20, 48 (1949).
[8] R. Bowling Barnes and C. J. Burton, J. App. Phys. 20, 286 (1949).

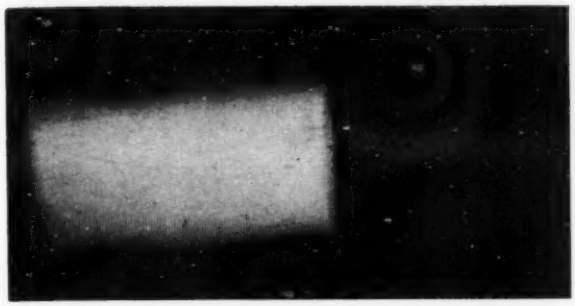

FIG. 3. Ultrasonic beam incident upon an aluminum plate at 0°.

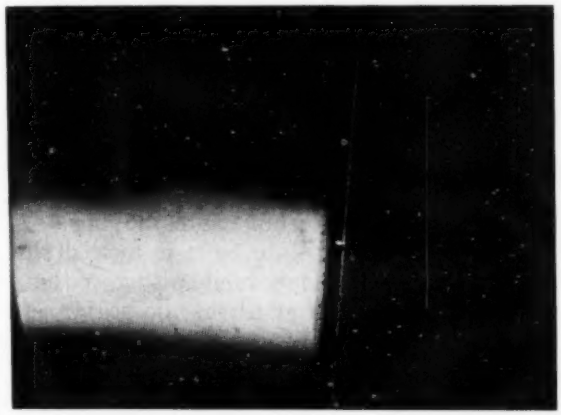

FIG. 4. Ultrasonic beam incident on an aluminum plate at 5°.

ous materials at discrete angles with respect to the ultrasonic beam was constructed and used during the course of this investigation. A schematic diagram and a photograph of the equipment taken directly from the preceding paper[8] are included in Figs. 1 and 2, respectively.

For precise visual measurement of angles and careful observation of the phenomena of transmission and reflection of ultrasonic waves it is highly desirable to employ a narrow beam or "pencil" of ultrasonic energy. Since, however, the wave-length of 5-megacycle ultrasonic vibrations in water is 0.3 mm, narrow beams of energy as employed in the present investigation are not very wide in terms of wave-lengths (*ca.* 25–30λ) and as a result diffraction effects are pronounced. This situation presents some difficulty in visual ultrasonic studies, and there appears to be no simple solution to the problem. To a limited extent the problem can be alleviated by using as narrow a beam as possible since when the beam width is halved the diffraction effects are condensed into one quarter the distance from the radiator. In the present work, an effort has been made to use a small source ($\frac{3}{8}$″ diameter), to mask with a suitable set of acoustically absorbent baffles the diffracted lobes which appear and to work as far as possible from the source consistent with the absorbing properties of the tank liquid. Under these conditions, diffraction effects are minimized and present no great problem in practical work. It should be noted in passing that diffraction effects of this nature are present in similar optical experiments although in this case the dimensions of the apparatus are so large compared with the wave-lengths of light that the effects become negligible.

Another problem of some difficulty is the construction of a tank of small dimensions which will at the same time be anechoic. In the past, various workers have gone to great lengths in an effort to circumvent this problem: experiments have been carried out in lakes which approximate an "infinite" medium; tanks of considerable size have been constructed with double walls, the inner wall being constructed of some material (e.g., rho-c rubber) whose acoustic impedance closely matches that of water and with the space between the walls filled with some highly absorbing material. In the present study, no extreme measures of this sort have been taken. The tank has been lined with rubber and several layers of copper window screen have been placed on the bottom and ends of the tank to prevent the formation of strong reflections. Some

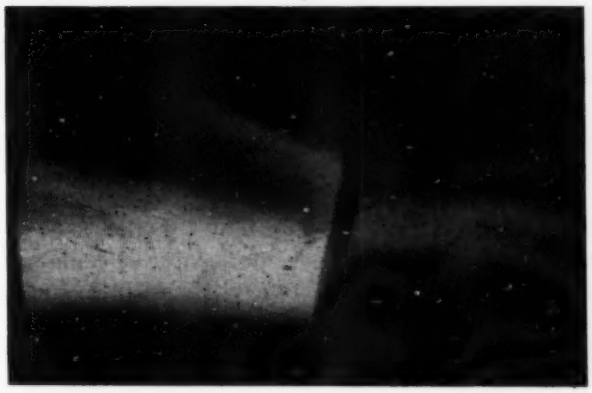

FIG. 5. Ultrasonic beam incident on an aluminum plate at 10°.

FIG. 6. Ultrasonic beam incident on an aluminum plate at 15°.

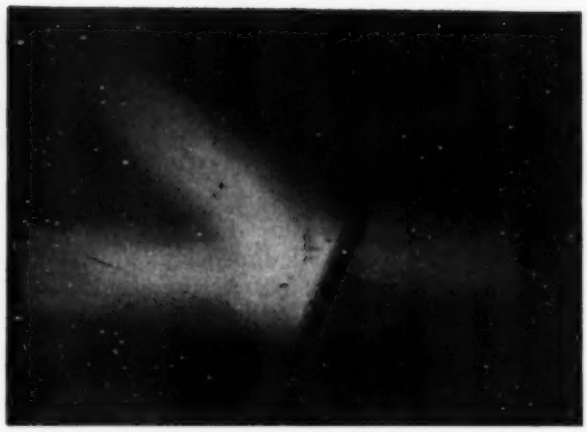

FIG. 7. Ultrasonic beam incident on an aluminum plate at 20°.

spurious reflections still exist, particularly those resulting from water wave reflection, although these have been relatively easy to identify and have not interfered with study of the primary plate transmission and reflection.

A series of photographs showing the nature of the transmission and reflection of an ultrasonic beam by a $\frac{1}{16}$-inch aluminum plate for angles of incidence between 0° and 45° at intervals of 5° is included as Figs. 3–12. A careful study of these photographs brings out several interesting features. The initial transmission is relatively small, decreases somewhat at 5° angle, increases through 10° and 15°, increases even more at 20°, decreases through 25° and becomes zero when the angle increases beyond approximately 30°. The reflections of the acoustic waves are also of interest: noticeable specular reflection is evident at 5° and 10°, with a second reflected beam making its appearance at 10°. Calculation indicates that the second reflection can be readily explained as a result of a partial reflection of the ultrasonic beam at the aluminum-water interface. The phenomenon is not unfamiliar in optics,

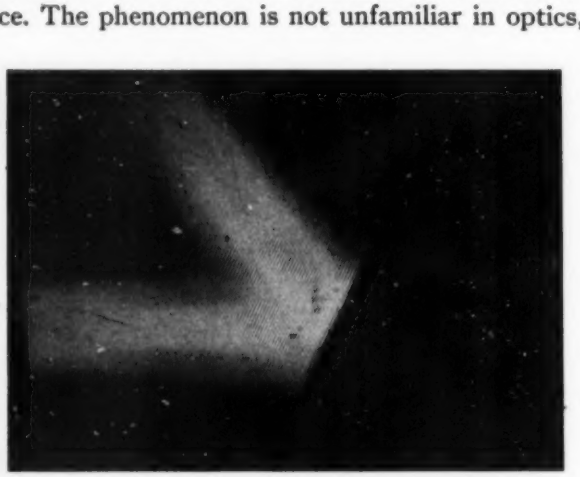

FIG. 8. Ultrasonic beam incident on an aluminum plate at 25°.

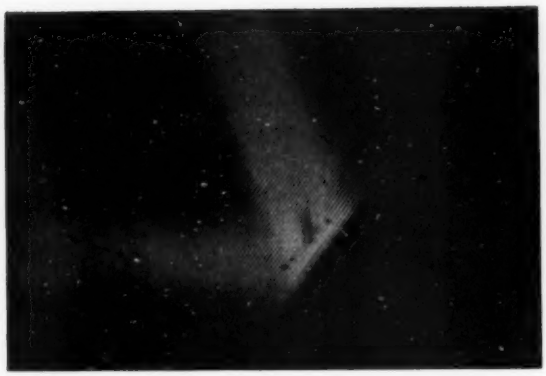

FIG. 9. Ultrasonic beam incident on an aluminum plate at 30°.

witness the mechanism of the Fabry-Perot interferometer, the Lummer-Gehrcke plate, etc. Actually, at intermediate angles as many as 7 reflected and an equal number of transmitted beams can be seen and photographed. It is extremely interesting to watch the "parade" of interference bands moving up and down the face of the plate as it is moved slowly through these regions. One should also note that the refraction in the case of an ultrasonic beam passing from a liquid to a metal is *away from the normal*, as opposed to the similar case in optics where light under these conditions is bent toward the normal. This situation arises, of course, because of the greater velocity of sound in a metal than in a liquid. The multiple-beam reflection disappears after reaching an angle of approximately 30° and strictly specular reflection then obtains.

As previously pointed out, these phenomena can be partially explained by analogy with familiar optical effects. On the other hand, many of the observed details do not appear to be explainable on this basis. This is understandable: light is a transverse wave phenomenon, whereas sound in fluids is propagated as a compressional wave and in a solid as either a compressional or shear wave, or both. It has been stated by other workers[9] in the field that as many as six different transmitted waves may result from the incidence of a single ultrasonic wave upon a thin plate of material, and there is considerable difficulty in separating observationally the longitudinal and transverse wave systems. Furthermore, there are at certain angles of incidence so-called "exchange waves" which are produced by the transverse wave, generated at the farther wall by a longitudinal wave in the plate, interfering in phase, on meeting the front wall and being reflected again, with a second entering ray. The extreme complexity of this situation, in addition to the fact that certain of the theoretical points are still highly speculative, makes it virtually impossible to analyze the problem completely with any degree of certainty. The photo-

[9] R. Bar and A. Walti, Helv. Phys. Acta **7**, 658 (1934).

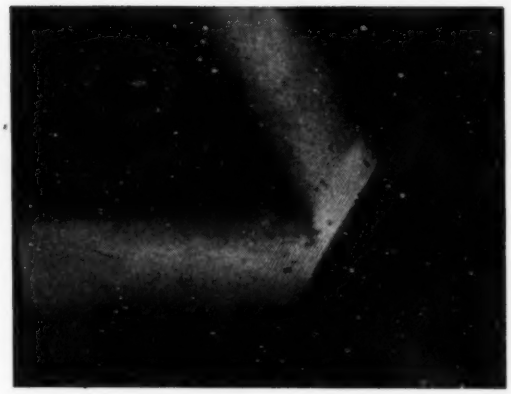

FIG. 10. Ultrasonic beam incident on an aluminum plate at 35°.

FIG. 11. Ultrasonic beam incident on an aluminum plate at 40°.

graphs themselves, simple as they appear at first glance, are in reality quite complex and do not clarify the problem of ultrasonic transmission through plates to the extent which we had initially hoped.

It is pertinent at this point to consider the mathematical relations which are required for calculation of the mechanical moduli from ultrasonic transmission data. The indices of refraction for dilatation and shear waves are given by

$$n_d=\frac{\sin\theta}{\sin\varphi_d}=\frac{V_l}{V_d}$$

$$n_s=\frac{\sin\theta}{\sin\varphi_s}=\frac{V_l}{V_s}$$

where

n_d = index of refraction for dilatation waves,
n_s = index of refraction for shear waves,
V_l = velocity of sound in the liquid,
V_d = velocity of dilatation waves,
V_s = velocity of shear waves.

The angles of total reflection θ_1 and θ_2 are given by the preceding equation when φ_d and φ_s become 90° and their sines become 1. Under these conditions

$$V_d=\frac{V_1}{\sin\theta_1}$$

$$V_s=\frac{V_1}{\sin\theta_2}.$$

The velocity of propagation of the shear waves is

$$V_s=\left(\frac{G}{\rho}\right)^{\frac{1}{2}}$$

where G = modulus of rigidity, ρ = density of plate. The velocity of the dilatation waves is

$$V_d=\left(\frac{(1-\mu)E}{\rho(1+\mu)(1-2\mu)}\right)^{\frac{1}{2}}$$

where E = Young's modulus, μ = Poisson's ratio. Poisson's ratio is expressed as

$$\mu=\frac{k^2-2}{2(k^2-1)}$$

where

$$k=V_d/V_s.$$

And the three constants are inter-related in the following way:

$$E=2(1+\mu)G.$$

The foregoing equations are of course strictly correct only when applied to homogeneous, isotropic media of infinite extent. These conditions are obviously not met in the case of a rolled sheet of metal, nor are they met in practically any experimental arrangement which can be set up. Since, however, it is impossible to specify the degree of inhomogeneity or the extent of non-isotropy, it appears legitimate to use these expressions as a first-order approximation to the "true" values of the dynamic moduli.

It has been shown[1] that characteristic transmission patterns are obtained when the acoustic energy transmitted through a plate is plotted as a function

FIG. 12. Ultrasonic beam incident on an aluminum plate at 45°.

of the angle of incidence of the energy. Figure 13[7] is an example of such a curve, showing the transmission through a $\frac{1}{16}$-inch sheet of aluminum. One notes that at 12° and 30° there is approximately zero transmission of acoustic energy. This is interpreted to mean that total reflection of the longitudinal waves which are excited in the plate occurs at approximately 12°. Beyond this point strong shear waves are set up in the plate, their intensity building up to a maximum and then receding to zero as the point of total shear reflection is reached. The periodic structure in both the longitudinal and shear portions of the curve is, according to one interpretation, a manifestation of the interference between the "exchange" waves and the incident longitudinal waves. There appear to be conflicting opinions[10] in this regard although reference to our earlier work[7] will show that these interference maxima can be used to determine the indices of refraction for dilatation and shear waves.

The angles of total reflection for the shear and dilatation waves can also be estimated with a fair degree of precision by inspection of the photographs in Figs. 14–16. It can be seen in Fig. 14 that there is zero transmission through the plate at this specific angle (11.5°). In Fig. 5, on the other hand, definite transmission is visible showing clearly the rapid building up of the shear wave intensity in the plate at an angle of 15°. From our previous discussion of the transmission curve (Fig. 13) it will be recalled that this corresponds to the region of total reflection for longitudinal waves. Thus, it is possible to estimate that this angle of total reflection for aluminum occurs at 11.5°. Similarly, the angle of total reflection for the shear waves is that angle beyond which all transmission ceases. From Figs. 15 and 16 it can be seen that this effect occurs at approximately 31.5°.

Using these data and substituting in the mathematical expressions recorded previously, the values

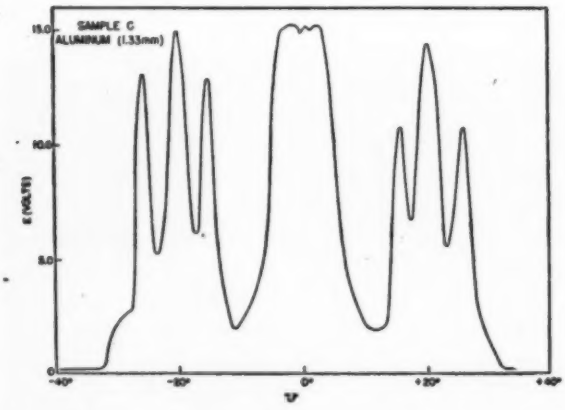

FIG. 13. Ultrasonic transmission pattern for $\frac{1}{16}$-inch aluminum plate, plotted electrically.

[10] J. Gotz, Akustische Zeits. **8**, 145 (1943).

TABLE I.†

	Electrical* recording	Visual recording	Published** value
Dilatational velocity (cm/sec.) V_d	7.05×10^5	7.0×10^5	
Shear velocity (cm/sec.) V_s	2.82×10^5	2.9×10^5	
Poisson's ratio	0.40	0.4	0.33
Shear modulus (dynes/cm²) G	2.14×10^{11}	2.3×10^{11}	$2.3–2.7\times10^{11}$
Young's modulus (dynes/cm²) E	6.73×10^{11}	6.4×10^{11}	$6.3–7.5\times10^{11}$

† Material—Aluminum sheet ($\frac{1}{16}$ inch); specific gravity—2.70.
* See reference 6.
** Ländolt-Boernstein tables.

of the longitudinal and shear velocities, the mechanical moduli and Poisson's ratio are readily obtained. The calculated values are recorded in Table I and are compared with those previously obtained using the electrical method of recording ultrasonic transmission data and with published handbook values. There is good agreement with handbook data save in the case of Poisson's ratio which is somewhat higher than the accepted value. However, in considering these data, it must be recognized that the calculations of mechanical properties have been made merely to show that rapid measurement of these properties is possible with simple equipment and with reasonable accuracy. Moreover, virtually all handbook data on mechanical properties are a result of static determinations, and dynamic measurements of the type shown here cannot be expected to be in perfect agreement with these values.

In conclusion, it is felt that the foregoing experiments are instructive for illustrating the complexity of the phenomena of ultrasonic plate transmission and reflection. In addition, the technique may be used to determine rapidly and with reasonable precision (±1°) the angles of total reflection for dilatation and shear waves. These values in turn may be used to calculate the dynamical values of Young's and the shear modulus with a precision limited mainly by the determination of the dilatational wave velocity (*ca.* ±8 percent). It is of course obvi-

FIG. 14. Ultrasonic beam incident on an aluminum plate at 11.5°, just at the critical angle for total reflection of dilatation waves. Note relatively high transmission in Figs. 4 and 5, for 10° and 15° respectively.

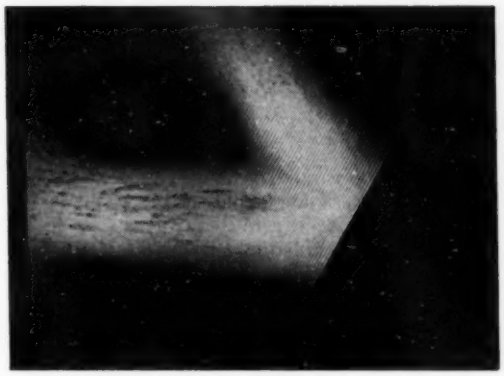

FIG. 15. Ultrasonic beam incident on an aluminum plate at 29.5°, just before reaching the angle of total reflection for shear waves.

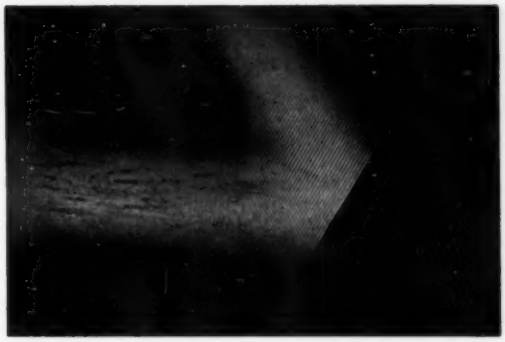

FIG. 16. Ultrasonic beam incident on an aluminum plate at 31.5°, just at the critical angle for shear wave reflections.

ous that the precision of the method as a technique for measuring mechanical moduli can be improved to almost any desired extent by devising either electrical or mechanical methods for improving the precision of angle measurement. However, these experiments have been carried out primarily to observe ultrasonic reflection phenomena by metal plates and the calculations of mechanical moduli have been made merely to illustrate the fact that the angles of total reflection for dilatational and shear waves can be observed easily and measured rapidly.

The High Current Carbon Arc and Its Mechanism

WOLFGANG FINKELNBURG
Engineer Research and Development Laboratories, Fort Belvoir, Virginia
(Received October 26, 1948)

The high current carbon arc with maximum crater brightness of 2000 candles per mm², axial intensity of 1.2×10^6 candles, power inputs up to 200 kw, light efficiencies exceeding 90 lumens per watt, radiation efficiency of 73 percent, crater temperatures up to 8000°K, and arc stream temperatures of the order of 12,000°K is the most powerful and one of the most interesting radiation sources known. The rising voltage characteristic, the anodic vapor stream which causes its excellent radiation properties, and the contracted arc stream distinguish the high current carbon arc from the well-known normal low current carbon arc. The properties of the contracted arc stream are discussed and a general theory of arc streams is presented. Starting from the anodic mechanism for the low current carbon arc, the anodic mechanism of the high current arc is developed. The rising voltage characteristic, the anodic vapor stream, and the high crater brightness are explained as a consequence of an abnormal voltage drop in front of the anode which increases with increased current density and is caused by a very rapid evaporation of the anodic material. The important role which the magnetic field of the arc current plays in the stabilization of all high current arcs is pointed out. The author believes that this unique high temperature arc has bearing on other arc and spark discharges, and also will play an important role in future developments of high temperature physics and chemistry.

INTRODUCTION

THE high current carbon arc, especially in the most common form of the high intensity arc or Beck arc, has, for more than three decades, provided the most powerful and brilliant light source in existence.[1] Only recently is this unique high temperature arc discharge finding a more general interest among physicists and chemists. In the first place, it is the only easily accessible source where the behavior of matter and its reactions can be studied at temperatures between 6000 and 12,000°K (plasma physics, high temperature chemistry, etc.). In the second place, this arc displays certain features unique in gas discharge physics and hence an understanding of its mechanism helps our general knowledge of the more complicated arc and spark processes. In this connection, the high current carbon arc exhibits (a) the first example of a discharge whose properties are essentially determined by its own magnetic field; (b) an arc stream where like in a metal the electric and heat conductivity are due to free electrons; (c) the first well studied case of a discharge whose mechanism is fundamentally influenced by mechanical vapor streams which are produced by the discharge itself. All these phenomena seem to play an important role in spark discharges.[2]

THE HIGH CURRENT CARBON ARC AS A NEW CLASS OF ARC DISCHARGE

Figure 1 is a schematic sketch and Fig. 2 a photograph of a fully developed high current carbon arc. Both figures display two important distinctions of the high current carbon arc from the normal low current carbon arc: (a) the deeply molded positive crater is filled with a streaming, brilliant, high temperature vapor ejected from the base of the crater which accounts for the high crater brightness, and (b) for currents above 100 amperes the arc stream appears to be contracted into a slender flamelike column (Fig. 3) with a current density more than ten times that of the low current arc stream. Further distinctive properties will be discussed later.

Ever since its discovery by H. Beck 35 years ago, the high current carbon arc has been operated with a cored positive carbon containing primarily cerium fluoride and oxide at anodic current densities of 100 to 400 amp./cm². This exceeds the current density of the usual low current carbon arc by a factor of 3 to 10.

It was generally assumed that this cerium compound core material was a necessary condition for the operation of the high intensity arc. The author found that an essentially similar phenomena could be obtained with many kinds of positive electrodes if operated at a sufficiently high anodic current density.[3] Indeed, we now believe the high current carbon arc should be recognized as a new form of an arc discharge which appears whenever sufficiently high current density causes vapor eruptions from the anode and, in certain cases, also from the cathode. The term *high current arc* has been introduced to distinguish this general type of discharge from the included special high intensity arc using cerium compounds in the core of the positive carbon. The vapor eruption from the positive carbon influences the arc mechanism in a way to be described later. It should be mentioned that the general shape

[1] Wolfgang Finkelnburg, *Der Hochstromkohlebogen* (Verlag Julius Springer, Heidelberg, 1948). English presentation (not quite as complete): *The High Current Carbon Arc* (FIAT Final Report 1052—PB-81644, Dept. of Commerce, Washington, D. C., 1947).

[2] Wolfgang Finkelnburg, Phys. Rev. **74**, 222 and 1475 (1948).

[3] Wolfgang Finkelnburg, Zeits. f. Physik **112**, 305 (1939).

of the arc, especially the direction of the anodic vapor flame and the arc stream as shown in Figs. 1 and 2, is independent of its orientation in space and does not result from the convection of hot vapors or gases. It is, rather, a general property of all high current arcs of more than 100 amperes that their specific shape is greatly influenced by their own magnetic fields.* We call them, therefore, field-stabilized in contrast to the wall-, electrode-, or convection-stabilized arcs.[4]

THE ANODIC PROPERTIES OF THE HIGH CURRENT CARBON ARC

Before we begin with the discussion of the discharge mechanism, it seems convenient to list the most important properties of high current carbon arcs, specifically those of the high intensity arc insofar as radiation properties are concerned. As has been mentioned, the high current carbon arc is operated with anodic current densities between 100 and 400 amp./cm² compared with approximately 40 amp./cm² for the low current arc. Evidence for a fundamental difference in the arc mechanism is the rising voltage characteristic of the high current arc, which is caused by an increase of the anode voltage drop with increased current. Partly because of this larger current density and partly because of the larger anode drop, there results a comparatively large consumption of the positive carbon ranging from 10 to about 50 inches per hour compared with the average of 3 to 4 inches per hour for the low current arc. The author was able to prove that the consumption in the high current carbon arc is due to evaporation, whereas in the low current arc it is predominantly due to chemical combustion.[5] This rapid evaporation of the positive carbon affects the core more than the harder carbon shell and thus causes the deeply molded crater. As a consequence of this evaporation and an additional heating effect, to be discussed later, the positive crater of the high current carbon arc in contrast to the low current arc is filled with brilliantly radiating vapors. These vapors, ejected from the carbon, form the anodic vapor "flame" mentioned above and indicated in Figs. 1 and 2. Whereas in the low current carbon arc only the incandescent electrode surface at a temperature of 4000°K or below is responsible for the radiation, in the high current arc the crater vapors at temperatures of 6000 to 8000°K are the main source of radiation. If these anodic vapors arise from compounds containing elements like cerium, which has closely spaced spectral lines throughout the visible region, copious amounts of white light are emitted. Crater brilliancies of arcs in practical application range from 500 to 1600 candles per mm² and brilliancies exceeding 2000 candles/mm² have been measured with experimental arcs. This brilliancy exceeds that of the surface of the sun (1600 candles/mm²). A total radiation as high as 6 kw per cm² crater cross section has been obtained. This total radiation is equal to that of the sun's surface and that of a blackbody at 5800°K. In further contrast to the low current carbon arc, the light efficiency is amazingly high. More than 70 percent of the total power input of high intensity arcs is emitted as radiation, and the light efficiencies (including the radiation of the anodic vapor flame) range from 55 to 90 lumens per watt compared with 23 lumens per watt for one of the best low current carbon arcs. Table I presents a summary of interesting data for a number of representative high current arcs in comparison with a low current arc.[6]

THE CONTRACTED HIGH CURRENT ARC STREAM AND ITS THEORY

Although all the properties of the high current carbon arc discussed so far are essentially anodic

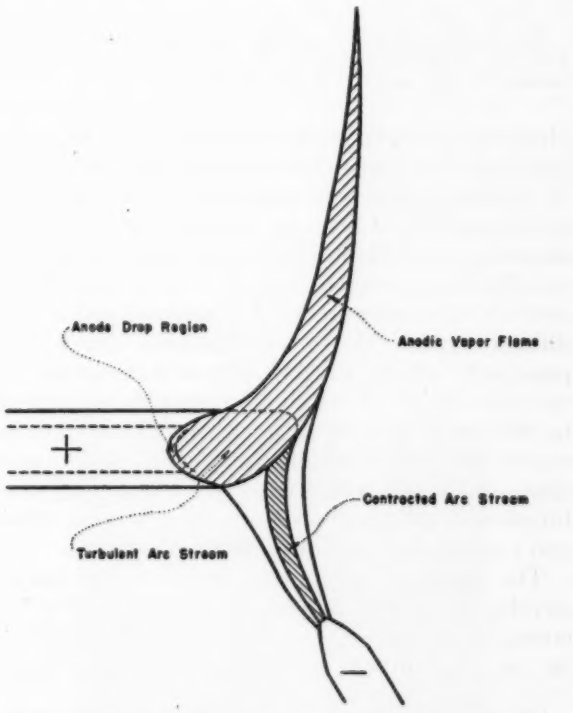

FIG. 1. Schematic sketch of the high current carbon arc with the different regions of theoretical importance.

* At currents above 500 amperes the magnetic field often causes a spiral curvature of the contracted arc stream and a very fast rotation of it round its foot point. This motion disturbs the arc considerably and prevents its steady operation. British searchlight investigators, not knowing what caused this unsteadiness, designated this arc state as the "exploded arc" (Symposium on Searchlights, Illuminating Eng. Soc., London 1948). All details of this phenomena, its theory, and methods of its prevention may be found on pages 11 and 127 of *The High Current Carbon Arc* (reference 1).

[4] Wolfgang Finkelnburg and K. H. Hoecker, Naturwiss. **33**, 55 (1946).

[5] Wolfgang Finkelnburg, *The High Current Carbon Arc* (reference 1), p. 77.

[6] Wolfgang Finkelnburg, Optik **2**, 149 (1947).

FIG. 2. Photograph of a high current carbon arc, burning with 185 amperes between water-cooled carbons (positive carbon 11 mm, negative 9 mm diameter).

phenomena, there exists a property common to all stationary arcs operated at sufficiently high current. At currents above 80 amperes the normal diffuse and expanded arc stream seems to contract into the above-mentioned bluish-white narrow column which we called contracted high current arc stream (Fig. 3), and which is surrounded by well defined zones of different color. Our investigations[7] indicate that practically all the current passes through the contracted column at the high current density of 1000 to 3000 amp./cm^2. It causes there a temperature in excess of 10,000°K which manifests itself by a strong continuous spectrum with superimposed widely broadened spectral lines of atomic nitrogen, oxygen, and carbon, but without any band spectra.

The theory of the contracted arc stream has been developed recently by Hoecker and the author[8] and corrected in details by Hoecker and Schulz.[9] In an arc stream of given diameter, according to Elenbaas,[10] the electric field strength E adjusts itself to such a value that the power input per cm arc length IE is compensated by the radial heat conduction and by radiation, S:

$$IE = 2r\pi k(dT/dr) + S,$$

where I is the current and k the coefficient of heat conduction of the arc stream plasma. In a molecular plasma like that of air, there are four processes contributing to the heat conductivity: the classical energy transfer by diffusing atoms (or molecules) k_a, that by free electrons k_e, that by molecules dissociating in the arc stream and recombining outside of it, thus conducting away the dissociation energy (k_d), and finally that by atoms ionized in the arc stream and carrying away the ionization energy. The contributions of each of these four factors have been computed for air as a function of the temperature by Hoecker.[8] His most interesting results are that below 7000°K the contribution of the free electrons, k_e, as well as that of the ionization, k_i, can be neglected, and in the region of 6800°K there exists a minimum of heat conductivity because the dissociation of the air is already complete and the ionization does not yet play an essential role. It is interesting that the measured axial temperatures of the low current arc stream in air agree fairly well with this theoretically computed temperature of 6800°K. At temperatures above 10,000°K, on the other hand, according to Hoecker's computation, the heat conductivity of the free electrons accounts for nearly all the heat conduction; all other processes can be neglected in comparison. The computation, carried through for a 200-ampere arc stream, resulted in an axial temperature of 11,000°K. Still higher temperatures are to be expected for higher currents. Consequently, the distinction between high and low current arc streams, introduced originally by the author for phenomenological reasons,[7] has found its clear physical meaning: The high current arc stream is distinguished from the low current arc stream by such a high a degree of ionization that the free electrons are predominantly responsible for the heat conductivity as well as for the optical behavior. The continuous spectrum results from the stopping of the free electrons in the electric field of the positive ions, and the few emitted lines of N, O, and C are widely broadened as a result of the electric fields of the surrounding electrons and ions.

Also the transition from low to high current arc stream seems understandable now. With increased current, the arc stream can either increase its diameter at the constant temperature of 6800°K, corresponding to the minimum of the total heat conductivity, or it can remain at constant diameter

[7] Wolfgang Finkelnburg, Zeits. f. Physik **116**, 214 (1940).
[8] K. H. Hoecker and W. Finkelnburg, Zeits. f. Naturforschung. 1, 305 and 382 (1946).
[9] K. H. Hoecker and P. Schulz, Zeits. f. Naturforschung. (1948), in press.
[10] W. Elenbaas, Physica 1, 673 (1934).

with increased temperature and consequently with increased electric conductivity. Above currents of approximately 100 amperes the second possibility seems to be more favorable for the energy balance because of the exponential increase of the electric conductivity with temperature.

THE ANODIC MECHANISM OF THE LOW CURRENT CARBON ARC

In order to understand the anodic mechanism of the high current carbon arc, we have to begin by discussing that of the low current arc. We will discuss several limiting cases and assume at first that the positive carbon does not emit positive ions even at its high temperature of the order of 4000°K. Immediately in front of the positive electrode the total current then is a pure electron current. The resulting negative space charge causes a potential drop immediately in front of the positive electrode surface. It is called the anode drop. By it, the electrons are accelerated and, by ionizing collisions with atmospheric gas molecules, produce the positive ions which diffuse in the electric field toward the cathode and which are necessary in order to compensate the negative space charge in the quasi-neutral arc stream.

Let it now be assumed that the positive carbon not only is able to emit positive ions at high temperature, as is the case to a certain extent for carbons containing metal compounds, but that its ion emissivity at 4000°K is sufficiently high to provide *all* the positive ions necessary for the quasineutral arc stream. Then no anode drop would be necessary for the production of ions, and we would find only a very small anode drop caused by the higher resistance of the "cooler" gas layer of 4000°K near the anode surface compared with the normal arc stream with a temperature of about 6000°K. But this state of the discharge is not a stable one. The incident energy of the bombarding electrons would not compensate for the cooling of the anode by radiation and heat conduction and hence its ion emission would fall below the necessary amount. But as this occurs the excess of electrons would cause an increasing space charge in front of the anode surface and thus increase the anode drop until a state of equilibrium is reached. Hence it follows that a certain part of the positive ions necessary is produced by ionization in the anode drop region while the rest is emitted by the anode surface, which is kept at sufficiently high temperature by the incident electrons accelerated by the anode drop. Thus a certain anode drop is a necessary consequence of the arc mechanism. The anode drop of the low current carbon arc using an ion-emitting, metal-cored positive carbon actually is only about 10 volts, whereas the anode drop for a pure homogeneous positive carbon without ion emission is approximately 30 volts.

THE ANODIC MECHANISM OF THE HIGH CURRENT CARBON ARC

Our interpretation of the anodic mechanism of the high current carbon arc, as distinguished from that of the low current arc, is based on one premise: the anodic vapor stream blows away the positive ions from the anode and thus, by an increased negative space charge, causes the increase of the voltage drop in front of the anode with all its consequences, and it is these consequences which are characteristic of the high current carbon arc.

To prove this premise, let us assume a low current carbon arc with the anodic mechanism described in the previous section. Now imagine a vapor stream of velocity u ejected from the anode surface through the anode drop into the arc stream. This velocity u has been found to be of the order of 10^3 cm/sec. and is vectorially added to the previously existing velocity of the positive ions and electrons in the arc stream. Since the velocity u is roughly 100 times that of the velocity of the positive ions in the arc stream due to the electric field, but only of the order of a few percent of the corresponding field velocity of the electrons, it is clearly evident that the positive ions are swept away from the region of the anode with a velocity essentially equal to u. A correspondingly larger number of ions, therefore, must be produced by ionization in front of the anode to compensate for those carried away by the vapor stream. This increased ionization requires an increased anodic voltage drop. How this is established, may readily be seen: While the positive ions are swept away out of the anode drop region with a greatly increased velocity, the electrons still travel out of the arc stream into the anode drop region and to the anode with a velocity which is not changed considerably by the vapor stream. Consequently, the negative excess space charge in the anode drop region, and with it the anode drop voltage, increases until sufficient ionization by the electrons, acceler-

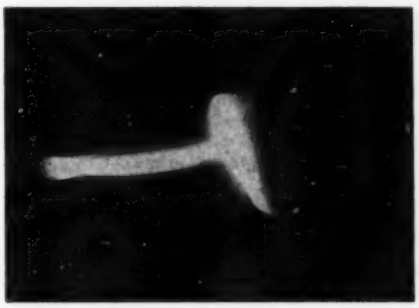

FIG. 3. Photograph of a contracted arc stream at 180 amperes, with positive carbon of so large diameter that current density is below the critical point; therefore no positive flame.

ated to greater velocity in the larger anode drop, establishes an equilibrium. In this stationary state not only enough ionization takes place in the anode drop region to compensate for the ions carried away by the vapor stream, but the electrons accelerated by the increased anode drop also carry to the anode enough energy in order to sustain by rapid vaporization the anodic vapor stream which is the dominant feature of the high current carbon arc and its anodic mechanism.

We can now explain the transition from the low current arc to the high current carbon arc as follows: By increasing the anodic current density we transfer more energy to the positive crater by the electrons accelerated in the normal anode drop than can be dissipated by heat radiation and conduction. Consequently, the anode surface begins to evaporate. As each vapor layer is pushed away normal to the anode surface by the new vapor being formed, as if ejected from a nozzle in the anode surface, the evaporation results in a vapor stream. The velocity of this stream is given by the amount of electrode material evaporating per second multiplied by the increase of the volume due to vaporization of the solid material and heating it to the temperature of 6000 to 8000°K.[11] The vapor stream causes an increase of the anodic voltage drop and thus of the electron energy transferred to the anode surface, so that in turn the evaporation rate increases until a state of equilibrium is reached. The energy transferred to the anode surface by the electrons is given by

$$W_a = i(V_a + \phi),$$

where i = electron current
V_a = anode drop
ϕ = work function

and is equal to the energy dissipated by heat conduction and radiation from the anode plus the amount consumed for the production of the high temperature vapor stream. The rising voltage characteristic of the high current arc in contrast to the falling characteristic of the low current arc is now explainable by our mechanism: an increase of the arc current causes an increase of the vapor stream velocity, and does so through increased vaporization. This in turn causes the increase of the voltage drop in front of the anode which has been measured by the author[7] and causes the rising voltage characteristic.

This anodic vapor stream is affected by a magnetic field because near the anode it forms a part of the arc stream (in Fig. 1 designated as the "turbulent arc stream") in which the electrons are mainly carrying the arc current. Any magnetic deflection of the electron stream in this region will be imparted to the vapor stream and cause an over-all change of its direction because the anode flame is a continuation of the vapor stream from the region of the turbulent

TABLE I. Characteristic data of high intensity arcs (see reference 6).

	Low current carbon arc	High intensity arcs in practical application					High load experimental arcs		
Diameter of positive carbon mm	10	7	7	9	16	24	20×30	7	10	13
Current amp.	30	50	75	75	200	450	1200	75	200	500
Arc voltage volts	55	40	55	50	78	90	100	65	85	~100
Arc wattage kw	1.65	2	4.1	3.75	15.6	40	120	4.9	17	~50
Consumption of positive carbon mm/h	100	320	1000	400	360	800	2000	2000	3000	4600
Brightness candles/mm²	188	550	1000	700	850	1100	1600	1400	1700	2000
Candle-power candles	9500	22,000	45,000	35,000	140,000	400,000	1,200,000	—	—	—
Total light output lumens	37,000	110,000	280,000	225,000	~900,000	~3,400,000	~10,000,000	—	—	—
Light efficiency lumens per watt	23	55	68	60	~60	~85	~90	—	—	—
Blackbody temperature of pos. crater T, °K	3800°	4300°	5000°	4800°	4500°	4700°	4800°	5500°	5400°	5800°

[11] Wolfgang Finkelnburg, Phys. Rev. 74, 1475 (1948).

arc stream. In the absence of external magnetic fields the anodic vapor stream is deflected in the direction of the bisector of the obtuse angle formed by the positive and negative carbons as shown in Figs. 1 and 2, by the magnetic field of the arc itself. This field, in trying to expand the arc current path, exerts an outwardly directed force on the arc stream, thus accounting for the curvature of the turbulent as well as the contracted arc stream as visible in Figs. 1 and 2. This outward curvature of both parts of the arc stream is greatest at their point of contact because here the inherent stiffness of both so-called "flames" is smallest, as has been discussed in detail elsewhere.[1,4] We believe that the magnetic field of the arc also helps in the contraction of the high current arc stream which may be regarded as consisting of parallel threads conducting the current in the same direction and therefore attracting each other. Because at currents greater than approximately 100 amp. these effects of the magnetic field predominate, as compared with other factors like convection currents of hot gases which influence any arc, Hoecker and the author[4] have designated high current arcs quite generally as "field-stabilized."

THE MECHANISM OF LIGHT EMISSION AND ITS DEPENDENCE ON THE ARC PARAMETERS

The influence of the arc parameters on the light emission by the high current carbon arc crater follows from the anodic mechanism discussed above. We have two ways of regarding the problem, though both in effect are equivalent. We may rely on the fact that in most arcs at higher pressures we have thermodynamic equilibrium throughout the arc except for a space extending a few electron mean-free-paths from each electrode. The emission of the crater vapors is then considered as thermal emission and the effect of the arc parameters on the vapor temperature in the crater must be determined. On the other hand, we may disregard the temperature concept and assume that excitation of the radiating crater vapors is entirely caused by electron collisions. The radiation depends in both cases on the density, average velocity, and collision frequency of the electrons, which in the first model are mainly responsible for the energy transfer and the heating of the crater vapors, while in the second model they directly excite the observed line emission of these vapors. We clearly distinguish two different regions of energy exchange (Fig. 1), the anode drop region, confined according to our calculations to about a few tenths of a millimeter, and the vapors in and in front of the crater through which the current passes. These vapors have been called the turbulent arc stream because of their rapid internal motion in contrast to the adjacent contracted arc stream in air (Figs. 1–3).

In the small anode drop region of a practical high current arc, the high power of approximately 6 kw is dissipated per cm^2 of crater surface. This is sufficient to transfer to *each* atom of the vapor stream an average energy of 8 ev, i.e., more than the ionization energy of the vapor atoms. The anode drop space is thus the main region where excitation and heating to a very high temperature take place. Consequently, the higher the anode drop for a given carbon the higher the crater brilliancy. High anode drop, according to our discussion of the anodic mechanism, is caused, however, by the high velocity of the anode vapor stream resulting from the very rapid evaporation of the positive carbon. Our theory then accounts for the empirical relation that high crater brilliancy requires high anode drop (i.e., high arc voltage) and high consumption of the positive carbons.

In addition to this energy exchange in the anode drop space, there is heating and excitation due to electron collisions in the turbulent arc stream within and immediately in front of the positive crater. As much as 2 to 3 kw is dissipated here per cm^2 cross section of the vapor stream. The influence of this arc region on the crater brilliancy, as measured from a point on the axis of the positive carbon, evidently depends on the current density in this region and on its thickness. This last point explains why the crater brilliancy depends on the crater depth. The same change, however, can be attained by influencing the thickness of the vapor layer in front of the crater, for instance by magnetic means.[12] There is an interesting possibility of distinguishing between the influence of the anode drop and that of the current density on the resulting crater brilliancy. Normally, by increasing the current the anode drop is increased together with the current density. We can influence the anode drop alone, however, either by changing the chemical composition of the positive carbon or, according to a recent investigation of the author,[13] by cooling the positive carbon immediately behind the burning end. By comparing carbons of different composition, or a cooled one with an uncooled one, at the same current density, we thus can study the influence of the anode drop on the radiation characteristics. The influence of the current density can be studied by adjusting the current of arcs with different carbon composition (or with a cooled and an uncooled positive carbon) so that they show the same anode drop. However, no systematic studies of the last sort have been made.

The effect of cooling the positive carbon, which has proved to be of great technical importance,[14] is

[12] Wolfgang Finkelnburg, *The High Current Carbon Arc* (reference 1), p. 88.

[13] Wolfgang Finkelnburg, J. Soc. Mot. Pict. Eng. **52**, (1949), in press.

[14] M. T. Jones and F. T. Bowditch, J. Soc. Mot. Pict. Eng. **52**, (1949), in press.

of considerable theoretical interest. The same positive carbon, when cooled up to the burning end, behaves quite differently from an uncooled carbon which is above incandescent temperature for at least one inch behind the burning end. This change in behavior seems to be caused by structural changes in the carbon due to the heating which are prevented by effective cooling. Such structural changes have been actually found by the author by x-ray diffraction studies which are as yet unpublished. The explanation of the cooling effect derived from these studies is as follows: Although the decrease of the anode drop which results from cooling causes a decrease of consumption and brilliancy for a *given* current density, the unsoftened solid structure of the cooled carbon allows operation at much higher load with consequent higher brilliancy. The overcompensating effect of an increased current density as against the moderate anode drop accounts for the surprisingly low positive carbon consumption per unit brilliancy.

THEORY OF THE ANODE DROP FOR LOW AND HIGH CURRENT CARBON ARCS

The foregoing sections have made evident the important role which the anode drop seems to play in the mechanism of the high current carbon arc. It must be possible, in principle at least, to develop a mathematical theory of the anode drop based on the anodic mechanism for the low and high current carbon arc described above. The difficulties encountered in this task seem to be of general interest in the field of arc physics.

An attempt at setting up a quantitative anode drop theory has been made by Heinzmann and the author.[15] Under certain simplifying conditions (monotonic change of the potential, velocity of electrons and ions given by the product of mobility and local field strength, field strength in the adjacent arc stream negligible compared with that in the anode drop, diffusion negligible) the distribution of potential, electric field strength, positive and negative space charge, and positive and negative current density could be calculated. In addition, a formula has been derived for the thickness of the anode drop region;

$$d = (12b^{-}V_a^2/\pi j)^{\frac{1}{3}},$$

in which b^{-} is the electron mobility, V_a the anode drop, and j the anodic current density.

The only difference between the systems of equations for the low current arc and the high current carbon arc is the introduction of the vapor stream velocity into the equations of the latter. We believe, by the way, that our anode drop theory for the low current arc, and with it our formula for its thickness, should apply for discharges quite generally, e.g., for glow discharges where it could be checked rather simply.

The anode drop itself could not be calculated but entered the equations as an empirical boundary condition. The reason for this deficiency of the theory is that we do not have an expression for the ionization in a high temperature discharge as a function of potential or local field strength in the regions of anode and cathode drop where at one end the thermal velocity of the electrons is larger than their field velocity, whereas the case is reversed at its other end.

OPEN PROBLEMS AND CONCLUSIONS FOR HIGH TEMPERATURE PHYSICS

In the last section we mentioned the unsolved problem of the ionization which is essentially a problem of high temperature physics. The same applies to the problem of how much the ionization potential of any atom or molecule is lowered owing to the presence of the large density of electrons and ions in the surrounding plasma, which, furthermore, influence the mobility of electrons and ions in a way not yet accurately known. A number of other details in the behavior of the contracted high current arc stream are still unexplained. How drastically the properties of an ordinary gas like air are changed by temperatures above 10,000°K becomes evident when it is realized that the electrical conductivity in such gases is of the same order as that of some metals and that the optical behavior (predominantly continuous emission and absorption) has little in common with that of "normal" gases.

These few remarks may suffice to indicate what interesting new insights may result from further study of the high temperature behavior of matter.[16] We think that the high current carbon arc is ideally suited for such studies because (a) it can be operated in any atmosphere at any desired pressure; (b) any kind of solid material can be introduced into the positive carbon core and thus vaporized and heated to plasma temperatures; (c) the arc in all its parts is easily accessible for detail studies.

The author is greatly indebted to Messrs. M. K. Laufer and G. W. Franks for their help in the final presentation of this paper.

[15] W. Finkelnburg and G. Heinzmann, unpublished description in reference 1.

[16] W. Finkelnburg, Naturwiss. **32**, 105 (1944); Die Chemische Technik **15**, 141 (1942).

Radio Doppler Investigation of Meteoric Heights and Velocities*

L. A. Manning, O. G. Villard, Jr., and A. M. Peterson
Electronics Research Laboratory of the Department of Electrical Engineering, Stanford University, Stanford, California
(Received October 1, 1948)

A continuous-wave Doppler method of determining meteoric velocities and heights was tested during the 1948 Perseid meteor shower. The accuracy of the technique was found to compare very favorably with optical and other radio methods. Doppler measurements have the advantage of relatively great sensitivity.

At the present time three methods are known for obtaining meteoric velocities directly from radio reflections returned by meteor ionization trails. The first to be described was that of Hey.[1] Working with the slow meteors of the Giacobini-Zinner shower of October 9, 1946, he obtained an average experimental velocity of 22.9 km/sec. with a standard deviation of 1.3 km/sec. The theoretical velocity is 23.7 km/sec. Hey's method consisted of a measurement of the change in range of the pulse echo as the meteor came in. He used a wave-length of five meters, and based his results upon 12 observations. Ellyett and Davies[2] have determined meteoric velocities by a second method based upon a measurement of the amplitude changes in pulse echoes from a filamentary meteor trail as the trail becomes extended in length. From observations of the Geminids, they deduced a mean velocity of 34.4 km/sec. with standard deviation of 2.8 km/sec., which they compared with Whipple's[3] photographic mean geocentric velocity of 34.7 km/sec. The third published method of determining meteoric velocities utilizes the Doppler frequency shift imparted to a continuous-wave signal by the motion of the reflecting cylinder of ionization.[4] The present paper describes a test of this technique. Velocities of the meteors of the August, 1948 Perseid shower were measured and the results found to be in good agreement with those obtained by optical means.

EXPERIMENTAL ARRANGEMENTS

For this test continuous-wave transmitters were operated on 23.1 and 30.66 Mc, as well as a pulse transmitter on 17 Mc for range determinations. The radiated power of the continuous-wave transmissions was of the order of magnitude of a kilowatt, as was the peak power of the 150-microsecond pulses. Both transmitting and receiving antennas consisted of half-wave dipoles, one-quarter wave high. The receivers were located in a small valley about four miles from the transmitters. For continuous-wave reception, the strength of the ground wave was kept within bounds by the sheltered nature of the valley, the orientation of the receiving dipoles, and by out-of-phase cancellation of a portion of the ground wave by means of a signal from directional antennas pointed towards the transmitters and so arranged as to be insensitive to signals arriving at angles much above the horizontal. Although a near-perfect ground-wave null is obtainable by the above methods, some ground-wave signal was allowed to leak through in order to beat with the reflected meteor signals.

The output of the continuous-wave receivers was recorded in two ways. A pair of d.c. amplifiers fed directly from the receiver second detectors were used to drive a two-channel direct-inking magnetic oscillograph. This oscillograph, employing Brush "pen-motor" units, has good response from zero to 120 c.p.s., and was operated at a paper speed of 125 mm/sec. It thus indicated all slow- and medium-speed variations in received field strength, and enabled a side-by-side comparison of the performance of the two channels. Doppler beats of higher frequency were recorded by means of a pair of magnetic tape recorders. For later analysis, these "whistles" were spliced out and played back into a standard two-channel oscillograph having a film speed of 48 inches per second. Thus, all Doppler shifts from zero to about 1500 c.p.s., were recorded for study.

The output of the pulse receiver was displayed on an "*A* 'scope" and slant ranges were read visually with the aid of ten-kilometer timing marks. The accuracy of range determination was of the order of plus or minus five kilometers for typical echoes.

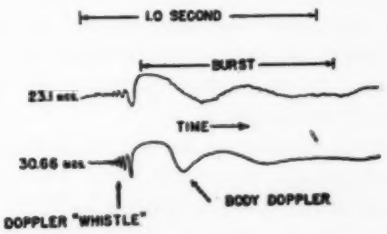

Fig. 1. Typical low frequency Doppler oscillogram.

* The research described in this paper was supported by the ONR and the US Army Signal Corps.

[1] J. S. Hey, "The Giacobinid meteor shower, 1946," Nature 159, 119–121 (1947).

[2] C. P. Ellyett and J. G. Davies, "Velocity of meteors measured by diffraction of radio waves from trails during formation," Nature 161, 596–597 (1948).

[3] F. L. Whipple, Proc. Am. Phil. Soc. 91, 189 (1947).

[4] L. A. Manning, "The theory of the radio detection of meteors," J. App. Phys. 19, 689–699 (1948).

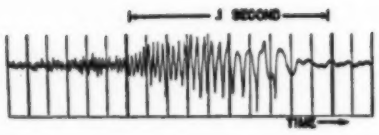

FIG. 2. Typical high frequency Doppler oscillogram.

Visual meteor observations were carried out on the night of the test with the aid of a team of observers facing the four quadrants of the sky. In order to obtain clear seeing conditions, the visual observations were made at an elevation of about 2000 feet in the Santa Cruz mountain range about three miles from the receivers. The bearings of the observed meteors were relayed to the receiving site by radio, for correlation with received echoes.

NATURE OF THE REFLECTIONS

The general nature of the phenomenon giving rise to Doppler meteor records has been described in a previous paper.[4] In Fig. 1 is shown a well-behaved penmotor field-strength record illustrating the principal effects of present interest. The "down whistle" with which the echo commences is associated with the component of meteoric velocity towards the observer, and falls below the audible range when the meteor arrives at that part of its straightline path nearest the observatory. It is this down whistle which is analyzed for velocity determination. In the case of large meteors, a non-linearity of the whistle-frequency *versus* time curve is sometimes observed at the lower whistle frequencies as a result of expansion of the ionization trail. The method of analysis of this non-linearity to determine trail size is a topic for a separate paper; in determining velocities, however, it is only necessary to restrict one-self to the high pitched part of the whistle in order to avoid inaccuracies due to the rate of trail formation.

A large increase in signal strength follows the down whistle because the incident wave is then reflected normal to the cylinder of ionization. It is this maximum signal, or "burst," that is seen on the *A*-'scope when determining range. The "body Doppler" which exists during the burst, or "body" of the echo, is of very much lower frequency than the down whistle and is of opposite sign, indicating a receding reflector. It is apparently caused by radial contraction of the fully formed column, again a subject for a separate report. The signal finally dies out when the ionization is dissipated by recombination and/or diffusion.

Figure 2 is a good example of a whistle oscillogram as recorded *via* the magnetic tape. This represents that portion of the down whistle which is too high in frequency to be recorded by the penmotor of Fig. 1. Unfortunately, the wave form is somewhat distorted by the magnetic recording process, but for the purposes of frequency analysis, wave form is non-essential. Although velocities of slow meteors can frequently be obtained from the portion of the whistle appearing on direct-inking oscillograms as in Fig. 1, fast meteors such as the Perseids have so high a rate of change of Doppler frequency with time, that an insufficient number of cycles is obtained within the frequency response of the instrument. It is then necessary to use magnetic tape oscillograms which show the higher whistle frequencies, as in Fig. 2.

THE PERSEID DATA

The measurements described here were made on the night of August 11–12, 1948, starting at approximately 10 p.m., P.S.T., and concluding at about 3 a.m. Visual observations were made in the interval from 11:50 p.m. to 2:30 a.m. That an abundant shower was in progress is apparent from the rate of visual detection presented in Fig. 3. A very large number of meteors was recorded by the radio equipment, but only those for which either an audible whistle or a visual azimuth were obtained, which could be definitely correlated with a pulse range, will be discussed here. A total of 138 whistles were catalogued, and visual azimuths were assigned to 36 of that number. Of the 138 meteors with recorded whistles, it was possible to determine velocities for 60; the remaining whistles were too weak to reduce or were otherwise inadequate. Only 14 of the 138 were analyzed on both 23 and 30 Mc, because of an unfortunate accident in the processing of the 23-Mc records. Velocities determined at the two frequencies were in satisfactory agreement. In the group of 36 whistles of meteors with azimuths, 20 were reducible to velocities.

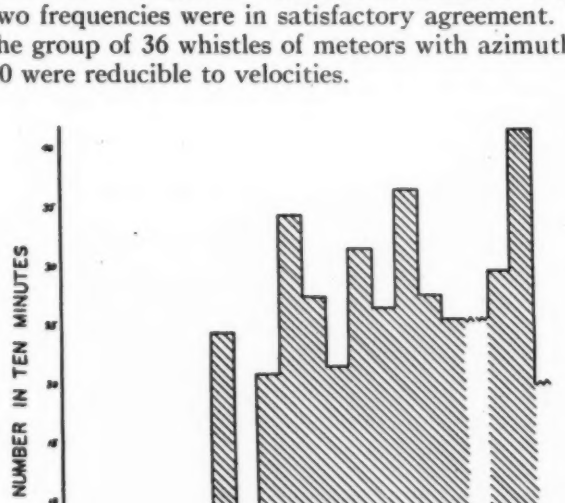

FIG. 3. Rate of visual detection—August 11, 1948.

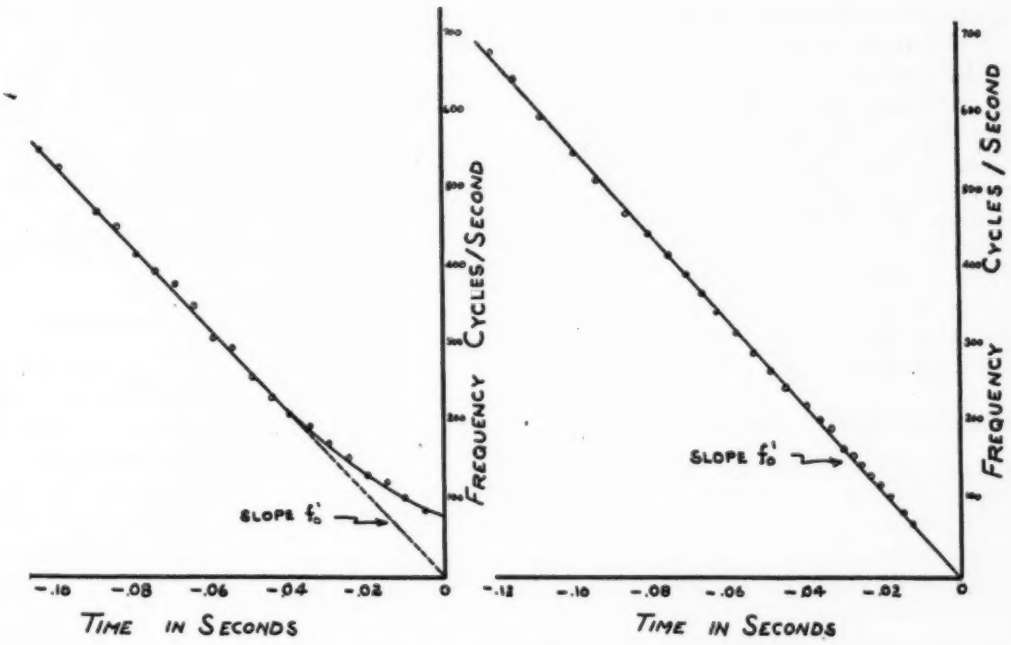

FIG. 4. Typical whistle-pitch *versus* time plots.

THE VELOCITIES

In a previous paper[4] one of the authors has shown that if the frequency, f, of a down whistle can be determined as a function of time, then a relation between the velocity and range of the meteor is specified. Denoting the slope with which the whistle pitch curve approaches zero frequency by f_0', the velocity of the meteor is

$$V=(-f_0'\lambda R/2)^{\frac{1}{2}},$$

where λ is the wave-length in the medium, and R is the pulse range. In Fig. 4 are shown two experimental curves of whistles pitch *versus* time. The first curve is linear and presents no questions in interpretation. The second curve is linear at the higher frequencies, but does not drop to zero as quickly as must the radial velocity of a meteor moving in a straight line with constant speed. The additional component of Doppler shift present at low frequencies is caused by radial expansion of the ionization column. The relation between this expansion and the Doppler shift has been studied theoretically and it has been found that the slope of a straight line drawn as shown will give the correct meteoric velocity despite diffusive expansion.

Velocities were determined for 60 meteors using the technique just described. In Fig. 5 is shown a plot of the resulting data, less six meteors for which either the range or the slope was in doubt. From the complete grouping of Fig. 5 it is possible to form two sub-groups with the aid of the visually-observed meteoric azimuths. Since we are now dealing exclusively with meteors producing a Doppler whistle, we know that the ray from the observatory is perpendicular to the direction of meteoric travel. The geometry of Fig. 6 must then apply to the reflection process. For each meteor it is possible to construct a line, L, seen in the views, which represents the intersection of a plane, P, perpendicular to the meteor path and containing the observatory, with a plane, M, at the height, h, of meteoric reflection. All shower meteors must produce their reflection when passing through L, so the difference between their observed azimuths A and the azimuth of the shower radiant Z must always be between 90 and 270°. In addition, a relation is specified between $(A-Z)$, h, and the pulse range R, if a meteor is to be along the line L determined by a radiant of elevation, H.

If we segregate those whistles having azimuth data into two groups as definitely Perseids or non-Perseids depending upon whether they satisfy the condition of the preceding paragraph, the distribution curves of Fig. 7 are obtained. It is strikingly evident in the case of these identified meteors that the Doppler velocities truly represent the shower speed. For the definite Perseids, the average velocity determined by the Doppler technique is 62.3 km/sec., with a standard deviation of 1.6 km/sec. There is one photographically-determined Perseid

FIG. 5. Meteor velocities—August 11, 1948 (average of Perseid meteors: 62.3 km/sec.).

velocity available for comparison.[5] Whipple found a velocity of 61.19 km/sec. with a deceleration of approximately 10 km/sec.^{2}. The accuracy of the photographic velocities is probably one percent.

In plotting Fig. 5 the definitely non-Perseid meteors of Fig. 7 have been cross-hatched with negative slope. It is seen that there remain at velocities between 56 and 69 km/sec. a single well-defined cluster. It is not at all certain that this cluster contains no non-Perseid meteors, but if we assume them to be Perseids we may compute an average velocity of 63.2 km/sec. with a standard deviation of 2.5 km/sec.

ACCURACY OF THE VELOCITIES

The accuracy of the Doppler velocities determined for the Perseids may be expressed in terms of standard deviation as a percentage of the average velocity. For the select group of definite Perseids, the accuracy figure becomes 2.6 percent, while for the group of presumed Perseids, the accuracy figure is 4.0 percent. By way of comparison, the same figure for Hey's determination of Giacobinid velocities[1] is 5.5 percent, and for Ellyett and Davies' Geminid observations, 8.1 percent. It should be pointed out that the Perseid meteors, because of their high speed, provide an especially severe test of radio methods of velocity determination in the following respect. If errors are to arise because of such effects as an "ionic bow wave"[6] lagging behind the particle, the effect should be most pronounced for fast meteors, because for them the ionization contours must have the greatest difficulty in keeping up.

Consideration of the cause of the observed scatter must begin with the accuracy of range determination. Ten-kilometer range marks were used, so that a perfect observer, reading range to within 10 km, would have an average error of 2.5 km. Assuming the actual accuracy of range determination to average 5 km, and assuming the average range as 130 km, the average error in the velocities would be about 1.2 km. As much or more error was probably introduced in finding the slope of the whistle pitch curves. Error of the latter sort can be reduced by careful attention to the quality of the whistle wave forms.

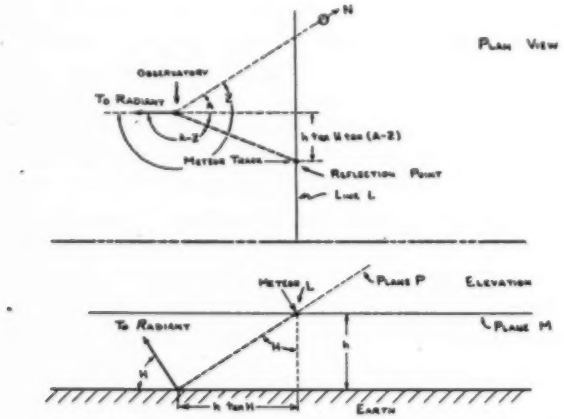

Fig. 6. Geometry of reflections from shower meteors.

HEIGHT OF THE PERSEIDS

Use can be made of the perpendicular nature of the reflection for meteors with whistles, to determine altitudes with good precision. Consideration of Fig. 6, or preferable its three-dimensional counterpart, shows that the pulse range may be represented in terms of its rectangular components as:

$$R^2=(h\tan H)^2+[h\tan H\tan(A-Z)]^2+h^2, \quad (2)$$

so that solving for the height h,

$$h=\frac{R}{(1+\tan^2 H\cdot\sec^2(A-Z))^{\frac{1}{2}}}. \quad (3)$$

The azimuth Z and elevation angle H of the meteoric radiant are very accurately known for shower meteors. The azimuth A can be observed visually, and if those meteors are chosen for which $\sec(A-Z)$ is near 1, the accuracy of the visual bearings becomes almost immaterial. Here is a simple method for determining the heights of shower meteors, in which the accuracy of height determination can approach that of pulse range measurement. As an illustration, suppose a meteor is chosen with an indicated angle $(A-Z)$ of 0°, and a radiant elevation H of 30°. Then a 5° error in azimuth determination introduces a mere 1/20 of one percent error in altitude.

In contrast, if altitude is determined by noting the pulse range and visual elevation of the meteor, errors in angle measurement are accentuated in computing heights, and uncertainty as to the part of

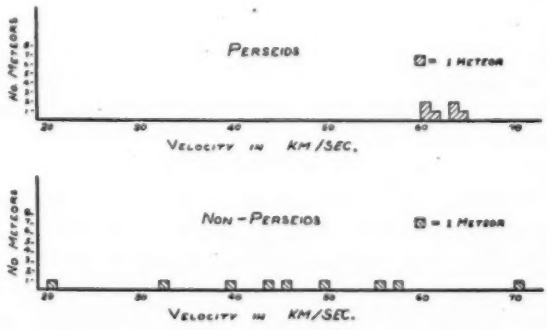

Fig. 7. Velocities of identified Perseid and non-Perseid meteors.

[5] F. L. Whipple, "Meteors and the earth's upper atmosphere," Rev. Mod. Phys. 15, 246–264 (1943).
[6] J. A. Pierce, "Ionization by meteoric bombardment," Phys. Rev. 71, 88–92 (1947).

the trail responsible for radio reflection introduces additional inaccuracy. In the present method the existence of the whistle is, of course, essential in that it guarantees the correctness of the geometrical construction.

From the group of meteors having both whistles and visually determined azimuths, thirteen Perseids were selected, and their heights as computed from Eq. (3) are presented in Fig. 8. The range is principally in the 90–110-km region, but a tendency for the altitudes to decrease with time is evident. Such a behavior might be anticipated from the behavior of the radiant, since its elevation angle was steadily increasing from a low value during the period of observation.

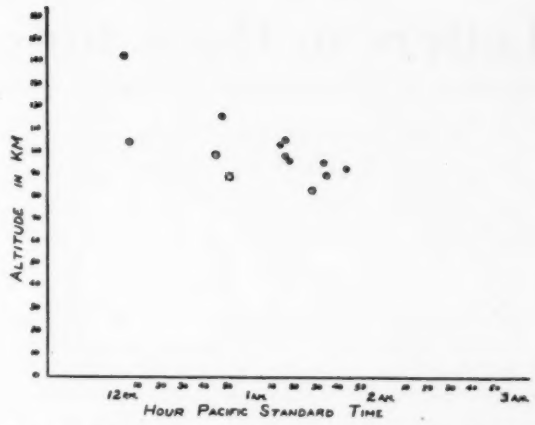

FIG. 8. Altitudes of identified Perseid shower meteors.

CONCLUSION

In this test the Doppler method of determining meteoric velocities has demonstrated itself to be remarkably accurate and reliable, deviations in measured velocities of only a few percent being observed for the Perseid meteors. The method has the advantage of relatively great sensitivity in comparison with other means of finding meteoric velocities. Study of shower meteors producing Doppler whistles has also been shown to be an exceptionally accurate way to obtain the altitudes of meteoric ionization columns.

Letters to the Editor

Improved Sectioning Technique for Electron Microscopy

RICHARD F. BAKER AND DANIEL C. PEASE

Departments of Experimental Medicine and Anatomy, School of Medicine, University of Southern California, Los Angeles, California

(April 27, 1949)

A METHOD of cutting sections with an ordinary microtome for electron microscopy was published about a year ago.[1] Sections 0.2-micron thick were described originally. Since that time, however, refinements have been made which make possible cutting and mounting 0.05-micron sections. It seems worth while to comment upon these improvements now, pending the publication of a fully detailed account.

Successful ultra-thin sectioning with a conventional microtome depends fundamentally upon three factors. 1. Suitable support must be given the specimen. 2. The knife must be sharp. 3. The adjustment of the knife tilt is critical. Also, since no commercial microtome is geared to cut fractional micron sections, an adujstment of the lever system is necessary. A very simply constructed adaptor for the Spencer No. 820 rotary microtome was described originally, designed so that the block advanced 0.1 micron with each cycle. The angle. of lever action has been further reduced so that 0.05-micron sections now can be cut.

FIG. 1. Electron micrograph of transverse section of doubly embedded hair. Two overlapping fibers are shown. Insert at higher magnification shows complex structure around edge. Section thickness apprx. 0.2 micron; 50 kv EMU microscope.

The original method of embedding specimens has not been significantly improved. After suitable fixation, biological material is dehydrated in a graded series of alcohols and then infiltrated with Parlodion (Mallinckrodt) dissolved in equal parts of ether and alcohol. The maximum concentration used is 12 percent. The Parlodion is then hardened in chloroform. It actually is precipitated as a very fine three-dimensional network. The block is then transferred to molten paraffin (m.p. 65°C). This replaces the chloroform and fills the interstices. When cooled the block is ready for sectioning. The Parlodion has sufficient elasticity and toughness to hold the thin sections together; the paraffin supplies the requisite hardness for cutting.

The knives used are standard microtome blades, and satisfactory edges can be obtained with hand sharpening. Our experience has been that a sharp edge will deteriorate within a week or so whether or not it is used. Accordingly it is efficient to employ automatic sharpening, and we now use a modified Fanz automatic knife sharpener (Arthur H. Thomas Company). Grinding with white rouge and soap is followed by polishing with plain water.

Very critical adjustment of the knife tilt is essential. The tilt is considered optimum when a section is cut with every cycle of the microtome. Outside the range of optimum tilt, which is not more than 0.5°, skipping and shredding of sections takes place. At the suggestion of Dr. J. Hillier an optical lever is used to measure the tilt. A horizontal light beam is reflected from the back facet of the knife bevel to a scale. The optimum knife position usually corresponds to an angle of about 2° between this facet and the block.

The most important new development concerns the handling of the sections once they are cut. Sections were originally flattened with needles by hand. By using dioxane the need for manual dexterity has been nearly eliminated. Sections are transferred from the knife edge with a small brush and placed on the surface of a drop of 50 percent dioxane in water. This begins to soften the embedding material, and surface tension starts to flatten the section. Then pure dioxane is added to the original drop. As the concentration increases, both the paraffin and the Parlodion are softened and partly removed. Interfacial tension completes the flattening. After the section has dried on the glass, residual paraffin is taken out with benzene, and finally the slide is dipped vertically into a 0.2 percent Parlodion in amyl acetate solution. This dissolves the original Parlodion and, after the slide has been drained, the sections are left embedded in a homogeneous film of controlled thickness. Screens are placed over the sections and the film floated off the slide onto a clean water surface. After picking up and drying they are ready for use.

Sectioning is normally done at room temperature. The sectioned bacteria of the cover illustration were produced in the manner described. In the case of very hard material, sectioning may be facilitated by lowering the temperature of the block. This increases the hardness of the embedding material so that more support is offered. We have used dry ice for this purpose, and with its aid have sectioned hair (Fig. 1), which otherwise ripped from the support. It is to be emphasized in this connection that although cooling always improves the cutting qualities of the block, it is never a substitute for a sharp knife properly tilted.

The thickness of the sections has been measured using a pre-shadowed replica technique. The sections were flattened on glass and completely extracted before being shadow-cast. Dow polystyrene latex spheres of known size were scattered around the section to serve as an internal standard. The thickness measured in this way is in satisfactory agreement with the value expected from the microtome calibration.

[1] D. C. Pease and R. F. Baker, Proc. Soc. Exper. Biol. and Med. **67**, 470–474 (1948).